

**INTENTIONAL SYSTEMS AND THE ARTIFICIAL
INTELLIGENCE (AI) HERMENEUTIC NETWORK:
AGENCY AND INTENTIONALITY IN EXPRESSIVE
COMPUTATIONAL SYSTEMS**

A Thesis
Presented to
The Academic Faculty

by

Jichen Zhu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Literature, Culture, and Communication

Georgia Institute of Technology
August 2009

**INTENTIONAL SYSTEMS AND THE ARTIFICIAL
INTELLIGENCE (AI) HERMENEUTIC NETWORK:
AGENCY AND INTENTIONALITY IN EXPRESSIVE
COMPUTATIONAL SYSTEMS**

Approved by:

Dr. D. Fox Harrell, Advisor
School of Literature, Culture, and
Communication
Georgia Institute of Technology

Dr. Jay D. Bolter
School of Literature, Culture, and
Communication
Georgia Institute of Technology

Dr. Kenneth J. Knoespel
School of Literature, Culture, and
Communication
Georgia Institute of Technology

Dr. Michael Mateas
Computer Science Department
University of California, Santa Cruz

Dr. Nick Montfort
Program in Writing and Humanistic
Studies
Massachusetts Institute of Technology

Date Approved: 05, July 2009

ACKNOWLEDGEMENTS

I am extremely fortunate to have the guidance and encouragement of my advisor, Fox Harrell. None of the work below would have been possible without his close attention and constant support, down to his offering painstaking comments. His vigor and open-mindedness have deeply influenced this dissertation and my approach to research.

I am also greatly grateful to members of my committee, Jay Bolter, Ken Knoespel, Michael Mateas, and Nick Montfort. Michael and Ken were among the first ones who introduced to me the concept of “expressive AI” and works in science studies respectively, which have joined to orient the foundation of this dissertation. Jay and Nick have provided me with valuable comments and advice at the times when I needed them the most. All of them have left firm stamps on this dissertation.

Especially, I would like to thank the following individuals who have offered their discussions, comments, and proof-reading at the crucial stages of my dissertation: Kurt Belgum, Sooraj Bhat, Jill Coffin, Steve Hodges, Madhur Khandelwal, Özge Samanci, and Geoff Thomas. Many thanks to Santi Ontañón, in particular, for sharing his computer science perspectives and for offering constant support throughout the entire process. Their support has made this dissertation significantly better.

I have also benefited greatly from the weekly discussions with the current and past members of the Imagination, Computation, and Expression (ICE) Lab, especially Kenny Chow, Ben Medler, Donna Sammender, Tonguc Sezen, Digdem Sezen, and Daniel Upton.

I am indebted to the diverse supportive communities in the Digital Media Program and Georgia Tech at large. I have benefited a great deal from my teachers,

particularly, Carl DiSalvo, Charles Isbell, Janet Murray, Nancy Nersessian, Ashwin Ram, Mark Riedl, Eugene Thacker, Bruce Walker, and Lisa Yaszek. I have learned as much from my fellow colleagues. At the risk of leaving out many names, I would like to thank Calvin Ashmore, Steven Dow, Clara Fernandez, Riccardo Fusaroli, Sergio Goldenberg, Mike Helms, David Jimison, Hartmut Koenitz, Hyun-Jean Lee, Brian Schrank, and Brian Sherwell. I also appreciate the administrative and technical support I received from Matthew McIntyre, and Grantley Bailey and Melanie Richard, respectively.

Many thanks to my friends here in Georgia Tech and also in Atlanta, many of whom have been mentioned above. Their company has provided balance to my life and made these five years some of the best years for me. I am equally grateful to my teachers, colleagues, and friends from Pittsburgh, Montreal, and Shanghai.

Finally, I would like to thank my family particularly my mother for their unconditional support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	x
SUMMARY	xii
I INTRODUCTION: INTENTIONAL SYSTEMS	1
1.1 The ELIZA Effect	3
1.2 Intentional Systems	5
1.3 Motivations	15
1.4 A Brief Account of <i>Memory, Reverie Machine</i>	17
1.5 Overview	18
II THEORETICAL FRAMEWORK	21
2.1 Perspectives on Computing	23
2.1.1 The Computer as a Stand-Alone Thinking Machine	24
2.1.2 The Computer as a Tool	26
2.1.3 The Computer as a Medium	29
2.1.4 Messages of Social Struggle	30
2.1.5 Narcissus' Mirror	33
2.2 System Intentionality	36
2.2.1 The Chinese Room Argument	38
2.2.2 The Intentional Stance	40
2.2.3 The Ghost outside the Machine: Social Perspectives	42
2.3 Reading Systems	48
2.3.1 Hermeneutics	48
2.3.2 Software Studies	54
2.3.3 Cognitive Semantics	57
2.3.4 AI and HCI Analyses	61

2.4	AI-based Interactive Narrative	72
2.4.1	Expressive AI Practice	73
2.4.2	Computational Narrative	74
III	THE AI HERMENEUTIC NETWORK	78
3.1	Intentional Systems as Text	81
3.1.1	What is <i>Text</i> ?	81
3.1.2	A Hermeneutic Framework	83
3.2	Constructing the Discursive Machine	85
3.2.1	Lessons from Alife: Discursive Strategies	87
3.2.2	The “Epidemic” of Intentional Vocabulary	90
3.2.3	The Code Level	95
3.2.4	The Presentation Level	100
3.2.5	Constructing the Literature	102
3.2.6	User’s Hermeneutic Reading	103
3.3	A Network Model	104
IV	THE SECRET OF HAPPINESS: A CLOSE READING OF THE <i>COPY- CAT</i> SYSTEM	107
4.1	Method	109
4.1.1	Principles of Hermeneutics	109
4.1.2	An Integrated Method	110
4.2	The Corpus and Potential Limitations	113
4.3	The Technical-Social-Cultural Context of <i>Copycat</i>	114
4.4	Content Analysis of the Technical Literature	119
4.4.1	The Two Aspects of <i>Copycat</i>	119
4.4.2	Intentional Vocabulary: Connecting the Two Languages	124
4.4.3	Leveraging the Two Languages of AI	133
4.5	Ideological Analysis	137
V	AGENCY PLAY AS A SCALE OF INTENTIONALITY	140

5.1	Scale of Intentionality and System Agency	141
5.1.1	Scale of Intentionality	141
5.1.2	System Agency	144
5.1.3	System Agency in Interactive Narratives	147
5.2	A Situated Approach to Agency	149
5.2.1	A Dance of Agency	150
5.2.2	Agency as Free Will	151
5.2.3	Agency as Resistance	153
5.2.4	Absence of Agency	153
5.3	Agency Play as an Expressive Tool	154
5.3.1	Agency Play	155
5.3.2	Agency Relationship	157
5.3.3	Agency Scope	159
5.3.4	Agency Dynamics	159
5.3.5	User Input Direction	160
VI	<i>MEMORY, REVERIE MACHINE: A CASE STUDY</i>	162
6.1	Motivation and Historical Context	164
6.1.1	Machine Memories, Reveries and Daydreams	166
6.1.2	Stream of Consciousness Literature and AI	167
6.1.3	Stream of Consciousness Literature and Cognitive Linguistics	170
6.1.4	Challenges of Engaging Legacy Forms	171
6.1.5	Related Works	172
6.2	Deployment of “Scale of Intentionality” and “Agency Play”	180
6.2.1	Main Indicators of System Intentionality	180
6.2.2	Examples of Various Levels of System Intentionality	182
6.3	Major Components	187
6.3.1	Dynamic Narration of Affect Using the Alloy Conceptual Blending Algorithm	189
6.3.2	The Emotional State Machine	190

6.3.3	Memory Structuring and Retrieval	191
VII	CONCLUSION AND FUTURE WORK	193
7.1	Revisiting the Major Arguments	194
7.1.1	The Formation of System Intentionality	194
7.1.2	Design Implications	196
7.2	Contributions	197
7.3	Future Directions	200
APPENDIX A	SAMPLE DATA FROM THE TECHNICAL LITERATURE ON <i>COPYCAT</i>	203
REFERENCES	216

LIST OF TABLES

1	Framework of Intentional Systems	14
2	Multiple Readings of <i>ELIZA</i>	15
3	AI Approach for Analyzing Systems	64
4	HCI Principles of Usability	69
5	Comparison between the AI and HCI Analyses	70
6	Timeline of Major AI Developments between 1980 and 1999	116
7	Usage of Intentional Vocabulary	127
8	Usage of Cognitive Faculties in Naming Functions and Structures . . .	128
9	Comparison of <i>Copycat</i> with humans and other forms of life	129
10	Indicators of “Scale of Intentionality” in <i>MRM</i>	182
11	Use of Intentional Vocabulary from paper “The Copycat Project: A Model of Mental Fluidity and Analogy-making.”	203

LIST OF FIGURES

1	George Lewis and His <i>Voyager</i> System	2
2	Original Transcript of <i>ELIZA</i> (Excerpt)	4
3	An Early Model of Tamagotchi	6
4	Four Characteristics of Intentional Systems (Prototype Model)	7
5	Screenshot of <i>AARON</i> 's Output	8
6	Screenshot of a Simulation of Braitenberg Vehicles	10
7	Original Transcript of <i>SHRDLU</i> (Excerpt)	11
8	The Main Fragment of the Antikythera Mechanism	24
9	Screenshots of Heider and Simmel's 1944 Film Experiments	37
10	Roomba Vacuum Cleaner with "Spotty Leopard" Costume	41
11	Mechanical Turkish Chess Player by Baron Wolfgang von Kempelen in 1769	42
12	CD Cover Art of Cope's 1997 Album <i>Classical Music Composed by Computer: Experiments in Musical Intelligence</i>	63
13	The AI Hermeneutic Network	80
14	Sample code of <i>ELIZA</i> from Norvig's Textbook <i>Paradigms of Artificial Intelligence Programming</i>	96
15	System Diagram of an Artificial Neural Network	99
16	A Partial Screenshot of the A.L.I.C.E Artificial Intelligence Foundation's Website	102
17	A Network Model of the AI Hermeneutic Network	105
18	Screenshot of a Java Re-implementation of <i>Copycat</i> by Scott Bolland	114
19	A Comparison between the Intentional and Technical Descriptions of <i>Copycat</i>	124
20	Source Code for a Bottom-up Scout Codelet	130
21	Source Code for One Kind of "Happiness"	133
22	Occurrences of Intentional Vocabulary in "The Copycat Project: A Model of Mental Fluidity and Analogy-making"	135

23	Density of Intentional Vocabulary in “The Copycat Project: A Model of Mental Fluidity and Analogy-making”	136
24	Lumière Brothers’ Early Film “Workers Leaving a Factory,” screened publicly in Paris in 1895	142
25	Dimensions of Agency	156
26	Architecture Diagram of <i>GRIOT</i>	188
27	An Annotated Example of <i>MRM</i> ’s Output	189

SUMMARY

Human interaction with technical artifacts is often mediated by treating them as if they are alive. We exclaim “my car doesn’t want to start,” or “my computer loves to crash.” Of increasing cultural importance are software systems designed explicitly to perform tasks and/or exhibit complex behaviors usually deemed as intentional human phenomena, including creating, improvising, and learning. Compared to the instrumental programs (e.g., *Adobe Photoshop*), these intentional systems (e.g., George Lewis’ musical system *Voyager*) seem to produce output that is “about” certain things in the world rather than the mere execution of algorithmic rules. This dissertation investigates such phenomena with two central research questions: (1) How is system intentionality formed? and (2) What are the design implications for building systems that utilize such intentionality as an expressive resource.

In the discourse of artificial intelligence (AI) practice, system intentionality is typically seen as a technical and ontological property of a computer program, emerging from its underlying algorithms and knowledge engineering. Distilling from the areas of hermeneutics, actor-network theory, cognitive semantics theory, and philosophy of mind, this dissertation proposes a humanistic and interpretive framework called the *AI hermeneutic network*. It accentuates that system intentionality is *narrated* and *interpreted* by its human creators and users in their socio-cultural settings. Special attention is paid to system authors’ discursive strategies, a constitutive component of AI, embedded in their source code and technical literature. The utility of the framework is demonstrated by a close analytical reading of a full-scale AI system, *Copycat*.

The theoretical discovery leads to new design strategies, namely *scale of intentionality* and *agency play*. They provide insights for using system intentionality and agency as expressive resources that can be used to convey meanings and express ideas. The fruits of these insights are illustrated by a stream of consciousness literature inspired interactive narrative project *Memory, Reverie Machine*, co-developed using Harrell's *GRIOT* system. It portrays a protagonist whose intentionality and agency vary dynamically in service of narrative needs.

CHAPTER I

INTRODUCTION: INTENTIONAL SYSTEMS

The labels in the exhibition said the usual: “Digital Print by Harold Cohen.” I called the curator and said, no, they are digital prints by AARON, a computer program written by Harold Cohen. The label was duly changed. But the bottom line still read: “Courtesy of the Artist.”

— Harold Cohen

I felt that I should be able to get the computer to sound good more or less on its own, so that someone listening to it says, “Who is that playing?” But if you get “What’s that?” instead, you have to go back to the drawing board.

— George Lewis [129]

Voyager is an improvisational interactive musical system developed by trombonist and composer George Lewis (Figure 1). The system receives and processes input from human jazz improvisers’ performances and generates music in real time. What differentiates *Voyager* from most other interactive digital music systems is neither the complex jazz patterns it generates, nor its ability to generate notes for many different instruments simultaneously, an ability that a human musician does not have. The difference lies in the status that Lewis grants to the system as an equal partner with other human improvisers. Lewis insists that his system is not designed to be “a musical instrument, but ... an independent improviser.” Even its designer cannot fully anticipate the music *Voyager* is going to play next, for it not only responds to other musicians’ playing, but also initiates independent behaviors that arise from



Figure 1: George Lewis and His *Voyager* System

its own internal processes. When left alone, *Voyager* can perform by itself without any human assistance. The musical dialogue between the system and jazz musicians performing with it, in Lewis’s mind, is a result of the “bi-directional transfer of intentionality through sound” [129].

Human interaction with technical artifacts is often mediated by treating them as if they are intentional, that is, with desires, beliefs, or dispositions about the real world. We exclaim “my car doesn’t *want* to start,” or “my computer *loves* to crash.” However, the phenomenon of casually describing artifacts as intentional is different from, for instance, regarding a computer program as an independent jazz improviser or painter *in its own right*. The latter exemplifies digital artifacts that are designed explicitly to support the narration and reading of its intentionality, as if their actions were *about* something in the world rather than just the execution of algorithmic rules. This chapter introduces the construct of *intentional systems* to distinguish such artifacts from the former case.

The concept of intentional systems is closely related to artificial intelligence (AI), cognitive science and arguably cybernetics. Many of the examples we will encounter in the dissertation are, in fact, AI projects. So why do I have to coin another term, instead of directly calling them AI or even cognitive science endeavors? The answer

has to do with the baggage that comes with AI. As a field whose past was tightly connected to military funding and cold war contexts [45], AI and computer science in large may impose ethical challenges to certain critical activists and artists [166]. Furthermore, AI and cognitive science are also communities of practice where their defining goals, valid methodologies, terminologies, and social interactions have been established and shared by its practitioners. Some practitioners, especially artists and designers, do not actively participate in these scientific/engineering communities, but are nevertheless constructing systems that engage the phenomenon of intentionality. Including Lewis, they would have been excluded from our discussion without broadening my scope. Hence, I propose the term intentional systems to downplay the boundary of the AI practice while focusing on its essential engagement of intentionality.

1.1 *The ELIZA Effect*

From the vantage point of the designers and engineers of digital systems and artifacts, the phrase “*the ELIZA Effect*” is used to describe the phenomenon that occurs when users perceive these systems as having “intrinsic qualities and abilities which the software controlling [them] cannot possibly achieve” or “assume that [the outputs of the system] reflect a greater causality than they actually do” [111]. The term is named after Joseph Weizenbaum’s renowned 1966 program *ELIZA*, which simulates a psychiatrist interviewing a patient. An original transcript of the program can be found in Figure 2. After the unexpected success of *ELIZA*, Weizenbaum was startled by “how quickly and how very deeply people conversing with DOCTOR¹ became emotionally involved with the computer and how unequivocally they anthropomorphize it” [212]. One of the famous anecdotes is that even Weizenbaum’s own secretary, whom Weizenbaum claims “surely knew it to be merely a computer program” after

¹DOCTOR is one of the applications of *ELIZA*.

Human: Men are all alike.

ELIZA: *IN WHAT WAY*

Human: They're always bugging us about something or other.

ELIZA: *CAN YOU THINK OF A SPECIFIC EXAMPLE*

Human: Well, my boyfriend made me come here.

ELIZA: *YOUR BOYFRIEND MADE YOU COME HERE*

Human: He says that I am depressed all the time.

ELIZA: *I AM SORRY TO HEAR YOU ARE DEPRESSED*

Figure 2: Original Transcript of *ELIZA* (Excerpt)

watching him working on the system, asked him to leave the room so that she could have a private conversation with the program.

As the creator of the program, Weizenbaum was aware of the simplicity of the underlying algorithm. Implemented on the basis of locating predefined keywords in user input sentences, *ELIZA* applies simple transformational operations to construct its output without further “understanding” of the conversation. Shortly after completing the program, Weizenbaum [211] openly acknowledged the user’s contribution to the illusion of a meaningful conversation with *ELIZA*. When interacting with the program, the user brings many assumptions from their real life experiences with human psychiatrists. For instance, if a patient gets the response “Tell me about boats” after telling her psychiatrist in real life “I went for a long boat ride,” she will not interpret it as the psychiatrist not having any knowledge about boats. More likely, she will assume that there is some purpose to this response in order to direct the subsequent conversation. Weizenbaum argues that it is the same assumption at work when the user interacts with *ELIZA*, despite its clearly limited dialogue capacity.

1.2 *Intentional Systems*

The ELIZA effect is a salient example of users’ attribution of intentionality to a system, a phenomenon closely related to what philosopher Daniel Dennett identifies as the *intentional stance* [38, 36]. According to Dennett, people generally adopt three strategies in order to predict the behaviors of humans, animals, artifacts, and even themselves [38, pp.298]. Some predictions are based on the laws of physics (the *physical stance*), while others are built on the assumption that artifacts follow how they are designed to behave (the *design stance*). For example, a cook can predict the effect of leaving the pot on the burner for too long using the physical stance, whereas he may adopt the design stance to make sense of his alarm clock without knowing much of its mechanical construction.

Sometimes, however, neither the physical stance nor the design stance is (completely) accessible. In this case, there is yet another stance one can adopt: the *intentional stance*. It entails treating artifacts, animals and humans as rational agents with beliefs and desires in order to further predict how they are going to behave [35]. For instance, we may not know exactly how a Roomba vacuum cleaner (Figure 10) is constructed or designed to traverse the entire room, but we can formulate our interpretations of its beliefs and desires based on observation and subsequently use them to predict its further actions.

An intentional system, Dennett continues, is any system that is “usefully and voluminously predictable from the intentional stance.” This definition aligns with the core of this dissertation, that is, the intentionality of the digital systems under study is *not* their intrinsic technical property, but rather is derived by the humans who create and use them. One issue remains, however.

In his dissertation, Seel [190] has shown that we can apply the intentional stance to almost all artifacts that we interact with, raising the question about the boundary of Dennett’s definition of intentional systems. To the majority of us, certain digital



Figure 3: An Early Model of Tamagotchi

artifacts (e.g., the humanoid robot *ASIMO*) afford intentional readings more easily than others. A robotic drummer, for instance, supports intentional readings more readily than a “hello world” program.

Various artifacts also support intentional readings in different ways. A doll does so, for example, differently than *ELIZA* or the *Voyager* system described at the beginning of this chapter. Construing dolls as intentional objects relies more on their humanoid forms than their relatively simple behaviors. Other artifacts display behaviors that we can relate to based on our own experience as intentional beings. The Japanese handheld tamagotchi (Figure 3), for example, mimics the behavior of a young animal pet who needs to be “fed,” “petted,” and “taken care of” by the owner pressing certain buttons. A small mistake while trying to satisfy the tamagotchi’s request could lead to its unexpected “death.” Similarly in Will Wright’s game *The Sims*, players are asked to take their virtual characters, called Sims, through daily activities such as sleeping, eating, bathing and dating. The Sims remarkably have a “free will” option. When this feature is turned on, the characters can move around autonomously and, without being controlled by the player, perform actions such as sleeping that are necessary for their own “well being.”

It is hence important to expand on Dennett’s term in order to emphasize certain

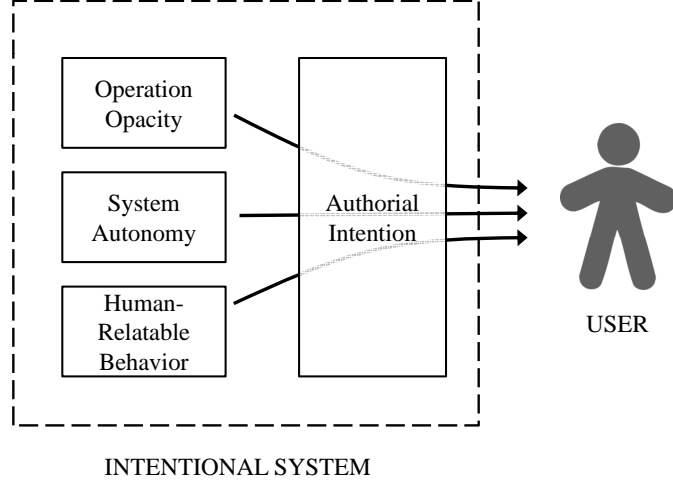


Figure 4: Four Characteristics of Intentional Systems (Prototype Model)

systems that seem to be more intentional than others, and to differentiate among various types of intentional readings by users. Instead of incorporating the entire range of artifacts to which the intentional stance can be applied, the rest of this section further develops the notion of intentional systems, this dissertation’s main area of inquiry.

First, the construct of intentional system follows a prototype model [179], where certain members of the category are more central than others. It means that the following characteristics may or may not all be present in practice, and the presence of each characteristic may vary by degree. I propose that intentional systems have four key characteristics: 1) operational opacity, 2) system autonomy, 3) human-relatable behaviors, and 4) authorial intention. Among them, 1) to 3) are mediated by 4) (Figure 4). The rest of this section elaborates these characteristics of intentional systems.

1. Operational Opacity

Operational opacity enhances users’ deployment of intentional stance. As Dennett’s theory has shown, when the operation through which a system derives its behaviors remains at least partially inaccessible through external observation,



Figure 5: Screenshot of *AARON*'s Output

users tend to attribute beliefs and desires to the system in order to make sense of these behaviors. A revealing example is Braitenberg Vehicles (Figure 6), a thought experiment constructed by cyberneticist Valentino Braitenberg [17]. The minimalistic vehicles move around in a virtual space in various manners that, “if we did not already know the principles behind the vehicles’ operation, we might call aggression, love, foresight and even optimism².” Those who look inside these vehicles, however, would find their control mechanisms extraordinarily simple, commonly in the form of “more light produces faster movement,” “less light produces slower movement,” and “darkness produces standstill.” Once their operational rules are unveiled, the user becomes less inclined to read these vehicles using an intentional stance, but rather based on these rules.

There are multiple ways that operational opacity can be achieved. Braitenberg vehicles exemplify the case where a system appears to be intentional when the underlying process controlling its behaviors is not directly available to the observer. Sometimes deliberately concealing or obfuscating a system’s underlying

²The quotation is from John Wiseman’s webpage “Braitenberg Vehicles” at <http://people.cs.uchicago.edu/~wiseman/vehicles/>

operation is its author’s strategy to construct its aura as an independent and intentional entity. For instance, digital artist Harold Cohen deliberately refuses to publish any detailed technical papers of his painting program *AARON* (Figure 5). Notably, within the community of computer science, AI practitioner Simon Colton adopts the same strategy for his recent system *The Painting Fool* [27], despite the field’s established tradition of unveiling the underlying operations.

In other times, even though a system author publishes her system in full details, the complexity of the algorithm makes it impractical to keep up with all the operations that are going on at any minute. In this case, an informed user and even sometimes the author herself still need to rely on intentional stance to a certain extent. Many AI projects fall into this category where operational opacity derives from its complexity. For example, neural network systems are notorious for their operational opacity. Its complex, multi-layered numerical weights make it nearly impossible for system authors to track their exact operations. In this case, it is much needed to read that the system is intentionally “learning.”

In general, when a digital artifact is intended to accomplish feats that we would casually describe as requiring human intelligence and intentionality, AI-level techniques that involve complex knowledge engineering and algorithmic approaches are generally utilized to sustain the intentional readings. An artifact without such complexity, on the other hand, may be able to accomplish such feats, but its behavior is typically predictable through a design or even physical stance once the user realizes its simple input/output pattern. For example, an electronic doll may express its “desire” to be “fed” by “crying” every time the baby bottle is removed from its mouth. However, the user can quickly notice the simple underlying binary control mechanism from its fixed response and therefore switch to a design or physical stance.

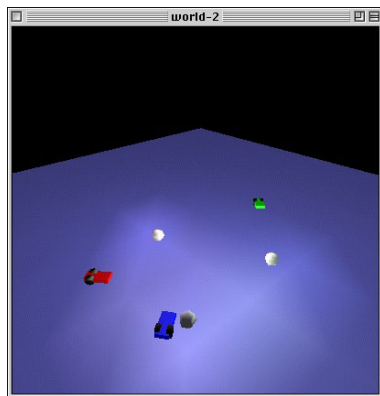


Figure 6: Screenshot of a Simulation of Braitenberg Vehicles

2. System Autonomy

An intentional system behaves autonomously, at least in certain aspects. There is a fundamental difference between a completely servile program (e.g., *Adobe Photoshop*, or *Ask Jeeves*) and one with some level of autonomy (e.g., the Microsoft Clippy³). Whereas the former passively obey user’s commands, the latter initiates interaction with user and takes actions on its own. Borrowing Lewis’s terminology, one is an “instrument” of the human user, while the other acts as an “independent player” [129]. *Voyager*’s intentionality partially relies on the fact that it does not completely follow the “commands” of other human jazz musicians, but rather displays independent behaviors from time to time.

System autonomy helps to explain the reasons why certain simple systems surpass many computationally complex ones as being more intentional. First, we will see an excerpt (Figure 7) from a demo of Terry Winograd’s acclaimed system *SHRDLU* [216], which moves various blocks on a simulated tabletop. Compared to the simplistic pattern-matching-based *ELIZA*, *SHRDLU* used many state-of-the-art natural language understanding algorithms to infer user input. Yet,

³From the vantage point of intentional systems, Clippy is a successful example, even though not a pleasant one. One of its problems, it seems, is that it is unfortunately placed in a context where a servile program is needed.

```

Human: Pick up a big red block.
SHRDLU: OK. (does it)
Human: Grasp the pyramid.
SHRDLU: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Human: Find a block which is taller than the one you
        are holding and put it into the box.
SHRDLU: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS
        TALLER THAN THE ONE I AM HOLDING.
SHRDLU: OK. (does it)
Human: What does the box contain?
SHRDLU: THE BLUE PYRAMID AND THE BLUE BLOCK.
Human: What is the pyramid supported by?
SHRDLU: THE BOX.

```

Figure 7: Original Transcript of *SHRDLU* (Excerpt)

when users watch the system following the exact command or faithfully reporting its implausibility, the system does not strike them as possessing beliefs or desires in the way *ELIZA* did. In fact, the design stance can sufficiently predict the former. By contrast, *ELIZA* appears to initiate and guide the conversation as an interlocutor with its own will. The (appearance of) autonomy helps to “elevate” a system from an instrument to an intentional agent.

SHRDLU is a special case because the simplicity of the domain in which it operates hides the complexity of the system’s operation and its potential autonomy. Despite the technical difficulty, the task of moving blocks on a flat surface is conceived as rather straightforward by an average human⁴. To be fair, the planning algorithm used in *SHRDLU* could have been read as more autonomous, if it were used in less-predictable domains such as airport management [91, pp.109-112].

3. Human-Relatable Behavior

The behavior of an intentional system has to be relatable to human users whose

⁴In the AI community, it is widely acknowledged that commonsense reasoning, a effortless task to human beings, is extremely difficult to realize computationally. This is one of the reasons why *SHRDLU*’s technical significance is easy to miss by many people who are familiar with the field of AI.

intentionality is directed towards the world. If a system fails to be predictable based on human-like beliefs and desires, it simply becomes incomprehensible. Human relatability does not exclude systems that mimic animal behaviors, such as the case of Tamagotchi, as most of them can be understood using human terms.

A good illustration of the importance of human-relatable behaviors is Joe Bates' influential OZ Project at Carnegie Mellon University. With the goal of creating "believable agents" using AI techniques [10], the project team did not follow the prior research on agents, which focused primarily on making the underlying algorithms more robust and efficient. Instead, Bates and others in the project introduced and highlighted "drama elements" and "interactivity" into these virtual agents, in order to make them behave in ways which are more human-relatable [136, 193].

4. Authorial Intention

The above three characteristics are so far concerned with *user's* perception of various aspects of the system, which is mediated and contextualized how the system author intends it to be. Here, I mean system author's intention of how her system should be regarded, whether intentional or not. We have already seen, based on different authorial intentions, how the system author can influence the perception of operational opacity by revealing/concealing the underlying process. For obvious reasons, authorial intention has remarkable impact on whether a system is designed to be (read as) autonomous and whether its behavior is human-relatable.

What I want to stress, however, is how authorial intention influence users' perceptions of a system, in addition to its design. Consider a java applet which generates colored pixels on a blank screen one line after another. After filling in

the entire display, it refreshes the screen and repeats this process, which renders completely different results. In addition, the source code of the applet is openly available to anyone. After observing the system’s behavior and carefully reading the code, a user may conclude that the system is quite autonomous (as it does not rely on any user command) but not very human-relatable. However, her opinion may be different after learning that the author of the program intends to use it as a commentary on the randomness of the stock market. With this new information, the user may see the system behavior as more human-relatable because now it is *about* something concrete in the world. In other words, the same code may be more or less intentional depending on what the author intends it to be.

A deeper reason, as elaborated in later chapters, is that intentional systems highlight the complex meaning exchange and negotiation between system authors’ narrations and users’ interpretations of intentionality. A system that is not designed to be read as intentional is not actively engaged in this process in the same way as intentional systems. Therefore, even though a car supports intentional readings under certain circumstances, it is not considered an intentional system in my framework.

In the practice of AI, authorial intention is implicit. Philip Agre calls critical attention to the role of narrated intentionality within AI by stating that the goal of AI is to build computational systems whose behaviors can be narrated using intentional terms, such as “planning” or “learning” [2]. As argued in greater detail in the later chapters, AI practitioners inevitably deploy rhetorical strategies in both their software and their publications, presentations, and interpersonal communications.

As mentioned earlier, my designation of intentional systems follows a prototype

Table 1: Framework of Intentional Systems (✓: High Degree; √: Low Degree)

	A Barbie doll	<i>ELIZA</i>	<i>Voyager</i>
operational opacity		✓	✓
system autonomy		✓	✓
human-relatability	✓	✓	✓
authorial intention	✓	✓	✓

model. For simplicity of representation, Table 1 and Table 2 use “high degree,” “low degree,” and absent to describe the level of intentionality in these systems. Table 1 provides reasons why *ELIZA* can be seen as more intentional than a Barbie doll since the former displays more human-relatable behavior as it maintains a natural language-based dialogue with the user. *ELIZA*’s underlying process, although simple, is more opaque than that of the doll. *Voyager*, on the other hand, is a more prototypically intentional system because it exhibits all four characteristics to a high degree.

For each characteristic, we should also take into consideration the specific social, cultural and technical context in which the system is situated. As shown in Table 2, the reading from a user of *ELIZA* in 1967 may be quite different from that of an average user today, even though the code of the system remains the same. Once a landmark in the history of AI and computing, *ELIZA* is considered a simple pattern matching program by the current AI community, a well-known phenomenon referred to as “the moving frontier of AI”⁵. College students are often given the assignment of re-implementing the program in their introductory level AI classes. In addition to its decreased complexity, *ELIZA*’s operational opacity also diminished with the society’s increasing level of procedural literacy brought about, in part, by the prevalence of digital products.

⁵The term of “moving frontier of AI” was first coined in the 1960s by Marvin Minsky, who called AI the study of computer problems that have not yet been solved [145]. This fast shifting standard of what constitutes AI is a fundamental phenomenon that is unique to the field.

Table 2: Multiple Readings of *ELIZA* (✓: high degree; ✓: low degree)

	A reading in 1967	A reading in 2009
operational opacity	✓	
system autonomy	✓	✓
human-relatability	✓	✓
authorial intention	✓	✓

1.3 *Motivations*

Today, we live in a world full of “smart,” complex, and intentional artifacts around us. We drive cars that can parallel park by themselves, play with robotic toy dinosaurs that have “personality” and do not always follow owners’ commands, and communicate with phones that suggest location-aware information before we ask for them. These latest consumer products may be reminiscent of certain familiar passages from our decade-old sci-fi books on AI. Unlike the rejections and heated debates that greeted the initial AI artifacts (fictional and actual) in the 1960s and 70s, the commercial success of their modern reincarnations suggest that society has come to terms with the existence of intentional but inanimate systems.

However, acceptance does not mean that we have gained better understanding of intentional systems. In fact, their increasing presence and cultural importance underscore the need to expand our insufficient knowledge of, among others, how these systems become intentional and how to design them. In order to do so, some people turn to technology for answers. They see machine intelligence and intentionality as merely technical properties of a system, proportional to the complexity of its underlying algorithmic work and knowledge representation. The techno-centric view of intentional systems is not a new approach, nor is it limited to the technologists; scholars from the areas of philosophies of mind, language, and AI, such as John Searle, assumed variations of this view in the debates of AI during the 1980s.

In contrast, the approach that I develop in this dissertation is to highlight the

complex meaning exchange process between system authors and users through intentional systems. In other words, this dissertation is about a humanistic view of AI, intentional systems, and digital technology. In order to do so, I turn to a long-standing field at the center of the humanities — hermeneutics.

For many people, the juxtaposition of AI and hermeneutics is quite an unusual combination. Hermeneutics as the study and art of interpretation dates back to the middle ages, where it was developed as a discipline for biblical studies. Later, under the work of Heidegger and Gadamer, hermeneutics was no longer limited to symbolic communication, and expanded into questions regarding human existence and humanism, including culture and arts (Chapter 2). In the age of digital technologies, hermeneutics in its broadest sense has become almost an icon of the humanistic tradition due to its deep concern for human cultural practices (e.g., literature, history, art, music, and philosophy) as well as even fundamental questions about human existence. As N. Katherine Hayles pointed out in a recent talk⁶, many scholars in the humanities fields share a common belief that *hermeneutic reading* is a privileged activity for humans, especially literary scholars. Machine reading, according to them, is insignificant by contrast because it is simply based on mechanical pattern searching without the rich human interpretations.

By bringing hermeneutics to the relatively young technological field of AI, I intend to propose an interpretive and humanistic view of AI as an alternative to the technocentric views. Part of my goal then is to tackle some of the questions that the technocentric views have so far failed to provide answers to, such as the how the Eliza effect occurs and the impact of the social and cultural dimensions of my four characteristics of intentional systems and their interpretation. My position as a critical theorist and AI practitioner allows me to “read” these intentional systems both at the code level

⁶Hayles’ public lecture titled “Minds, Machines, and Media” on January 15th, 2009 at the 2008-09 Georgia Tech School of Literature, Communication, and Culture’s Distinguished Speaker Series.

and at the discourse level, an approach, which has just emerged in the nascent field of software studies.

1.4 A Brief Account of Memory, Reverie Machine

Memory, Reverie Machine (MRM) is a text-based interactive narrative system that is designed to oscillate between a prototypical intentional system and a more servile one. Whereas more detailed discussions of the system can be found in Chapter 6, this section intends to paint a broad picture of the system, especially from the vantage point of users' experience.

The stories generated by *MRM* are about a robot character, Ales, who is jointly controlled by the user and the system. When the story starts, Ales functions as the user's avatar; he follows and responds to the user's commands completely and transparently. As the story develops, however, Ales's behaviors and processes become increasingly opaque to the user: he recalls past memories and daydreams from time to time without the user's permission; a coherent human-relatable personality and emotional state of the character may also emerge, depending on how the user interacts. In addition, the narration of the story shifts from an objective, third person point of view to a more subjective and personal voice. An excerpt of a transcript below illustrates *MRM*'s behavior at this point of the story.

(the color of the kitchen door today should be [yellow/blue])

> BLUE

(;) (ales said to himself, adjusting his visual scan system.)

(the room where he had his first encounter of tune-up and oil change)

(had similar depressing entryways.)

(the oil change left a sickly feeling in his gut;)

(he would rust like the tin man before enduring another.)

(it amazes him that no one else complained about these things;)

(they were too busy with the kitchen;)

(he should probably start his daily routine now.)

Compared to the onset of the story, the excerpt contains Ales’s memories of past events. Such memories, reveries, and daydreams are not directly controllable by the user, and the process by which they are retrieved remains inaccessible. The user, however, can have impact on Ales’s emotional state. Choosing “blue” over “yellow” in this case sets the character in a “melancholy” mood, which causes his surroundings to *appear* “depressing” from Ales’s subjective point of view.

As the story goes on, the system becomes even more intentional. It may direct Ales to ignore certain commands issued by the user and eventually takes complete control of Ales. At this late stage of the story, to sustain a design stance (i.e., the system, like most interactive narrative systems, is designed to follow the user’s command) is nearly impossible.

Several readings are available to make sense of the system. The user could either regard the *MRM* system as intentional for overseeing the story and its narration, whereas Ales is an avatar jointly controlled by the system and the user. Alternatively, the user could identify Ales and the system as the same entity, which gradually gains its intentionality. Regardless of which reading the user adopts, the main concept behind *MRM* is that the degree to which a system is intentional is a potential expressive mechanism in the design of computational systems.

1.5 Overview

This dissertation is organized around two central research questions: 1) How is system intentionality narrated and interpreted, and 2) Given an answer to (1), what are the design implications for building systems that utilize such intentionality as an expressive resource? This chapter has already introduced intentional systems —

a collection of computational systems, in both software and hardware forms, that exhibit complex behaviors usually seen as the territory of intentional human activities such as planning, learning, and creating. Chapter 2 offers an interdisciplinary theoretical framework drawn from hermeneutics, philosophy of mind, actor-network theory, and cognitive semantics. These different theories and approaches are synthesized into four major issues of close relevance to intentional systems. First, I provide an overview of the different historical perspectives on computing in order to situate the role of intentional systems in a broader context of computing practice. Next, the core concept of *system intentionality* is examined from different positions ranging from philosophical debates about AI to social theories that attribute any human-like agency to artifacts. Third, various possible approaches for reading intentional systems are introduced, followed by a brief review of existing interactive narrative works that influenced the design and construction of my co-authored narrative-based intentional system *Memory, Reverie Machine* (described in Chapter 6).

Chapters 3 and 4 are devoted to the first research question, i.e. how we can analyze intentional systems function and interpretation in a social context. Chapter 3 develops a novel theoretical framework — the *AI hermeneutic network* — to understand the formation of system intentionality. Compared with existing methods of analyzing computer systems in AI and human computer interaction (HCI), this framework emphasizes system authors’ discursive strategies for *narrating* system intentionality at both the code and discourse levels, as well as focusing on users’ interpretative readings of system intentionality in a broad social context. Chapter 4 exemplifies the utility of the AI hermeneutic network with a close reading of a real and well-known AI system, *Copycat*.

Building on these analytical frameworks and results, the following two chapters focus on their design implications for intentional systems, the second research question. Chapter 5 raises the design concept called *scale of intentionality*, first introduced in

Zhu and Harrell [220], which states that system intentionality can be manipulated to convey meanings. Accordingly, it proposes a socially-situated design model of *agency play* first articulated in Harrell and Zhu [84] that dissects agency in interactive narrative works into various dimensions and suggests different strategies to manipulate each dimension as an expressive resource. Chapter 6 introduces our own interactive narrative system, *Memory, Reverie Machine*, as a case study for applying the above design theory in expressive intentional systems. Finally, this dissertation ends with conclusions and suggestions for future research directions.

CHAPTER II

THEORETICAL FRAMEWORK

The machine’s danger to the society is not from the machine itself but form what man makes of it.

— Norbert Wiener, *The Human Use of Human Beings* [213], 1950

The intentional systems we have seen so far are simultaneously connected to social, cultural, and technical foundations. Their multiple situated connections impose unexpected challenges for applying the existing methods directly from any established disciplines. On one hand, traditional computer science analyses tend to be preoccupied with issues such as computational efficiency, task effectiveness, and generalizability of particular systems to other systems. These analyses often lead to the techno-centric view that the phenomenon of intentionality displayed by these systems is purely a result of the complexity of the underlying algorithm and/or the expressivity of the knowledge representation. This approach hence puts social and cultural issues aside, if acknowledging them at all. The fields of social sciences and humanities, on the other hand, have developed many powerful theoretical and methodological tools to study technology in broader social and cultural contexts. However, a majority of scholars from these fields do not have the technical means needed for unpacking these AI-based systems. Consequently, many choose to blackbox the systems’ underlying operation and the related literature generated, such as technical papers and interviews. This approach is also inadequate because, as shown in Chapter 3 and 6, the code and literature of intentional systems deeply connects to how their authors encode and narrate intentionality.

Recent years have witnessed an emerging trend of more interdisciplinary and integrated approaches that bridge the gaps between the techno-science sphere and the humanistic fields. For instance, the research area of human computer interaction (HCI) and the rising research area of software studies both push the limits of their originating fields of computer science and media studies respectively and reach over to the other side of the gap. In line with this development, this chapter presents an interdisciplinary framework for both close analysis of the *formation of intentionality* in intentional systems and correspondent *design strategies* for these systems. It is a crucial step because many of the systems studied in the later chapters have only been looked at inside the domain of computer science.

My theoretical framework is organized around major questions that arise in understanding the nature, function, and interpretation of intentional systems, rather than a survey of fields. These questions are: 1) What is the role of computational technology in our society? 2) Can (digital) artifacts have intentionality? If so, in what senses? 3) What existing framework of “reading” can help us to read intentional systems? and, 4) What approaches are useful to design intentional systems in the domain of interactive narrative. In addressing these questions, I will synthesize the related major fields and approaches, highlighting the ones that I feel shed the most light.

The rest of the chapter is organized around these four questions. Section 2.1 presents three major perspectives on computers/machines in relation to humans in the development of computing. It is important to go over this history because it helps to answer the question of how intentional systems integrate into society. Section 2.2 takes on the topic of system intentionality, a concept central to the systems of interest in this study. It challenges the traditional notion of intentionality based on the closely related debates in philosophy of mind/AI, as well as theoretical developments in other areas such as activity theory, actor-network theory, and distributed

cognition theories. Subsequently, a variety of theories and methods useful for reading intentional systems are introduced in Section 2.3. The hermeneutic tradition of text interpretation, emerging software studies approaches for reading computer programs, and cognitive semantics methods to analyze non-literary expressions described in this section are the foundation of the new method proposed in Chapter 3. From the area of cognitive semantics, conceptual blending theory is also of particular relevance for producing intentional systems, as illustrated by the computational narrative system *Memory, Reverie Machine* that I developed with Fox Harrell (Chapter 6). Related traditional and computational narrative antecedents are presented in Section 2.4.

2.1 Perspectives on Computing

In the history of modern computing, different visions of the relationship between humans and computers have competed for dominance. One of the ways to summarize the development and adoption of digital computers is through the different major perspectives that emerged since the mid-twentieth century. It is important to trace this history, for each of these perspectives represents a different way to position intentional systems in our society. Are they autonomous machines with their own beliefs and desires that might one day replace humans? Are they tools we can use? Or, perhaps, are they a medium through which we can convey meanings and express ourselves? The answer to these questions will determine how we see and interact with these digital systems.

As the historical development of computing has revealed, the debates about the computer's ontological status as human has gradually given way to the question of how society can benefit from these systems, either as tools or as comprising a medium. This observation is important for situating my approach and practice as described at the end of this section.



Figure 8: The Main Fragment of the Antikythera Mechanism, dated approximately to 150—100 B.C.

2.1.1 The Computer as a Stand-Alone Thinking Machine

“Can machines think?” Pioneers of computing proposed this daring question in the early twentieth century and generated heated debates across the different sectors of society accompanied by such vastly polarized responses as fear, curiosity, and excitement. It is closely related to the question of machine intentionality because, as many would argue in the following decades, intentionality has been considered to be a prerequisite of thinking and intelligence [188, 189]. On one hand, this question challenged many widely accepted civil and intellectual norms within fields such as philosophy, theology, and biology. On the other hand, it connected to the long-standing dream of building self-operating machinery, a tradition that can be traced to the Antikythera mechanism (see Figure 8), a device to calculate astronomical positions in Hellenistic Greece. It also resonates with the golem figure created from clay in traditional Jewish folklore, which has since become a symbol of “automata.”

A typical response that reassures the conservatives is the classic Lady Lovelace styled denial. Having worked with Charles Babbage and translated materials on his analytical engine — one of the earliest mechanical general-purpose computers — Lady Lovelace famously asserted that computers cannot originate anything since all they do

is to follow pre-defined instructions step by step. To the followers of Lady Lovelace, any linkage between the instruction-following machine and the action of thinking is out of the question.

This conventional position, however, was acutely challenged by the booming field of computer science and engineering, in particular artificial intelligence (AI) and cybernetics. Even though Alan Turing [202] tactically rephrased the question into “can machines *appear to think*” to avoid the controversies generated by the original one, the trajectory of the fields in the next few decades clearly downplayed the difference. After all, Turing clearly predicted in the same article that “[at] the end of the century, the uses of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.”

As a direct challenge to the Lady Lovelace argument, the AI dream of building an intelligent machine that can think has dominated the research areas of cybernetics and AI for decades — the former attempted to “model brains in computer hardware” through feedback-loop-based control scheme whereas the latter “sought to mimic minds in software” [45, pp.239]. Built on the assumption that thinking is information processing, the early AI community made a series of major breakthroughs, including Newell and Simon’s *General Problem Solver* [159], Daniel G. Bobrow’s *STUDENT* program [12], Weizenbaum’s *ELIZA* program [211], and Terry Winograd’s *SHDLU* [216]. These developments, part of what was later called “Good Old-Fashioned Artificial Intelligence” (GOFAI) [85] or “classic AI,” are built on Newell and Simon’s influential “*Physical Symbol Systems Hypothesis*”, which states that “A physical symbol system has the necessary and sufficient means for general intelligent action” [160].

More importantly, Paul N. Edwards [45] points out that another reason why AI received so much public attention and government/industrial funding is that both AI and computing were crucial to Cold War social and cultural contexts because they were consequently recognized as metaphors and political icons during that period of

time. For instance, John McCarthy’s concept of time-sharing not only provided “the right subjective environment for AI work” so that researchers could work interactively with the computer, but also greatly strengthened the link between AI and the military Command and Control scheme (pp.256-59).

The stand-alone AI dream was not confined to research labs and government funding agencies. It also ignited the public’s imagination and manifested itself through various popular media, especially science fiction literature and its related cultural productions. The AI-based systems and characters depicted in these fictional worlds are often based partially on imaginary and recognized goals and state-of-the-art of existing AI research. For instance, Michael Mateas [137] points out that the renowned AI system *HAL 9000* in Kubrick’s film *2001: A Space Odyssey* [114] “both captures preexisting intellectual currents that were already operating within the field of AI, and serves as an influential touchstone that had a profound impact on individual AI practitioners and on the aspirations of the field.”

The AI dream was constructed collaboratively by its researchers, government / military institutions, and popular media. This complex phenomenon deeply influenced the course of computers’ development. Unfortunately, the rapid development quickly filled the field with unrealistic ambitions and exploded expectations in what became known as the AI winter, which eventually led to the field’s shift to less ambitious and more practical focuses.

2.1.2 The Computer as a Tool

A parallel, but less radical, image of a computer is that it is not intelligent by itself, but is a useful tool for easing the tasks of humans. Different from the AI dream of a stand-alone thinking machine, this perspective underscores its instrumental utilities in service of the human. The systems following this perspective utilize the computer’s ability to organize and distribute information in order to help their users to accomplish

tasks varying from word processing to distance learning.

A well-known example of the tool-based computer system is Vannevar Bush’s conceptual design of the Memex. As the twentieth century witnessed an unprecedented explosion of information, a more practical challenge facing society than AI was how to better organize, distribute and make use of these data. In 1945, Bush [22] published on his vision of the Memex machine — an electromechanical information organizational device that helps a researcher to browse through information in its library and record her own trials of thoughts or follow other people’s through associative “traces” and notes between different contents.¹ The microfiche-reading device replaces the traditional mechanical way of indexing with a method similar to the associative nature of human cognition, in order to better assist human intelligence. The vision of Memex has influenced the development of computers in many aspects. It was a precursor to many new technologies such as hypertext, speech recognition, and World Wide Web, directly influencing many later pioneers of computing such as J. C. R. Licklider, Douglas Engelbart and Ted Nelson.

During the 1960s and 1970s, the notion of building computers to provide servile work to human was almost an antithesis of the much more ambitious vision of stand-alone AI. Even today, after the AI community shifted further away from the original AI dream to solve more specific tasks, the research agendas of HCI and AI are still mostly independent from one another, and at times oppositional. In-between the two perspectives, are systems developed as a tool to augment human intelligence.

Licklider’s concept of *man-computer symbiosis* [130], for example, envisioned collaboration between human and computer to “greatly improve the thinking process.”

¹An interesting parallel exists between Bush’s invention and his life-long career. The Memex’s major contribution was not intended to produce more knowledge by itself. Instead, it would organize existing information to facilitate further research and innovation. Despite his role as an established researcher, Bush is widely remembered as an organizer who helped to establish the close connection between the military, government, and academia. The so-called iron triangle has incubated a vast amount of scientific and technical advancement and contributed to America’s political power.

In his idea of such a symbiosis he wrote, “[m]en will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking” [130]. On one hand, the “man”-computer symbiosis is different from the notion of the cyborg and the AI dreams in that Licklider’s vision was built on the differences between human and machine capabilities, as opposed to constructing the almighty human/machine that replaces one, the other, or both.² On the other hand, the machine in the symbiosis is also a “buddy” that assists the human to get things done, and therefore can be seen as a bridge between the two sometimes conflicting perspectives.³

To a certain degree, Douglas Engelbart’s work is a continuation of Licklider’s. Engelbart saw computers as tools to boost humankind’s capability for coping with complex, urgent problems. Pursuing his goal of augmenting human intelligence, Engelbart undertook the task of helping human users to gain “more rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble” [48]. The inventions Engelbart was directly involved in, including the computer mouse, word processing programs, and early stage graphic user interface (GUI), were all important milestones in the development of human computer interaction (HCI) paradigms.

²The term *cyborg* originated in the 1960s in relation to the field of cybernetics. It refers to a hybrid creature between human and machine, the machine typically being a form of computer and information technology.

³Historically, Licklider is often associated with AI because of his directorship at the Information Processing Techniques Office (IPTO) — part of the early Defense Advanced Research Projects Agency (DARPA) organization — turned the IPTO into an incubator of early AI projects by providing them with large amounts of funding.

2.1.3 The Computer as a Medium

A third perspective on computing is that the computer is a communication medium, a potentially expressive one. This perspective is certainly in dialogue with McLuhan's broad notion of a medium [141] as an extension of human capabilities. In response to the limitation of the media studies approach, McLuhan raised the difference between the "medium" and the "content" and argued that existing studies narrowly focused on the content at the price of the medium. In this regard, he proclaimed famously that "the medium is the message." Using the electric light as an example — spelling out the name of a brand or lighting certain activities are at the content level, since neither usage could exist without the electric light. But "it is the medium that shapes and controls the scale and form of human association and action [I]t is only too typical that the 'content' of any medium blinds us to the character of the medium." According to McLuhan, the fact that electric light had escaped the attention of the media scholars unless it was used to spell out an advertisement or a company brand provided evidence that people had "fail[ed] to study media." Instead, it is the medium or technology itself that should be studied prior to the content it carries, if any.

As a technological object, the social impacts of computers easily comprise a McLuhanian "medium," but a computer is not just a medium in the same sense as the electric light. Artists, designers, writers, and scholars have noticed the potential of computers as a communication medium precisely because its "content" can incorporate other media. Ted Nelson [157] and Alan Kay [109] were among the first ones who adopted the viewpoint that the computer is a medium. They both encouraged people to understand computers and to become literate in programming to master the medium. Contradictory to McLuhan's statement, Kay claims that "[t]he essence of a medium is very much dependent on the way messages are embedded, changed, and viewed. Although digital computers were originally designed to do arithmetic computation, the ability to simulate the details of any descriptive model means that

the computer, viewed as a medium itself, can be all other media if the embedding and viewing methods are sufficiently well provided.” The later inventions such email and World Wide Web all have reinforced the popularity of this view.

With the advancement of digital technologies, computers are increasingly regarded as (potentially) an expressive medium. However, Janet Murray [154] rightfully reminded the overly optimistic people the difference between an artistic medium and the technology that supports it, with a reference to the historic moment of film in the Grand Café at Paris in 1895. She wrote “[t]he legend of the Paris café is satisfying to us now because it falsely conflates the arrival of the representational technology with the arrival of the artistic medium, as if the manufacture of the camera alone gave us the movies” (pp.66). This observation led to a variety of work that sought to define the formal properties and aesthetics of the digital medium/media [154, 134]. These studies are often done in the style of Greenburg’s art criticism [74], in the sense that they intend to identify the “intrinsic” and identifying characteristics of the digital medium that distinguishes it from other ones. In spite of the variety of perspectives proposed, the current digital expressive practices loosely fit into what may be called the “procedural” approach [154, 134, 135, 14, 210] and the “metamedium” [109] approach. The former deems the computer’s ability to generate different outcomes as the essence of the medium, whereas the latter highlights its capability to incorporate different media forms. In the domain of narrative, many interactive and generative narrative systems overall are examples of into the first approach, while many electronic literature (eLit) works are close to the second. Detailed discussion of computer-based narrative works can be found later in Section 2.4.

2.1.4 Messages of Social Struggle

The previous sections have sketched out three major perspectives: the stand-alone AI view, the instrumental view and the media view, in the early history of computing.

Today, their descendants still occupy substantial areas such as AI, HCI and digital media and arts respectively. So far, I have mainly described these perspectives as independent of each other. In reality, they are under the constant influence of each other.

For instance, the question of whether machines can think has faded away from society’s attention as a driving controversy. On the one hand, the AI view has conceded to the instrumental view of computing as the field of AI has shifted its primary focus to solving practical problems that can benefit humans. Even for those still working on the general problem of thinking, the problem is often framed not as that of creating a stand-alone entity, but rather as creating as cognitive prostheses for humans [58]. Yet, the field has also drawn inspirations from other perspectives. One example is Mateas’ Expressive AI [135] that engages the medium view by turning AI into an expressive medium in its own right.

The rest of Section 2.1 shifts the focus slightly from the practical functions of computers to their *symbolic* roles in society. The arrival of computer technologies has destabilized many established boundaries, such as the one between human and machine, and the one between natural and constructed. The nuanced position of the computer lends itself to rhetorical and iconic strength as an advocate for social struggle and change. Compared with the previous approach of computer as a medium, some theorists and artists use computers directly as the message itself.

“Though both are bound in the spiral dance, I would rather be a cyborg than a goddess,” Donna Haraway [76] famously declared. Haraway constructed a socialist feminist mythology of her cyborg — a cybernetic organism that oscillates between organism and machine, life and death, social reality and fictional creature — in order to offset the rigid and homogeneous society. In her cyborg world, “people are not afraid of their joint kinship with animals and machines, not afraid of permanently partial identities and contradictory standpoints.” Of particular interest to this study

are the design principles that Haraway proposes on both material and ideological grounds. What she calls “the informatics of domination” takes on the issues of “comfortable old hierarchical dominations” of gender, race, and class by underscoring that the concepts of purity and unity are as imaginative as the mythical cyborg itself.

If Haraway’s manifesto of cyborgs engages the metaphor of information technology as a fragmented hybrid, George Lewis [129] instantiates it through the actual construction of computer systems. The *Voyager* system, mentioned in the beginning of Chapter 1, is deliberately constructed as a rebel against the prevailing hierarchical interaction model of human-leader-computer-follower. Instead, the system embodies a non-hierarchical *multidominance* concept, initially brought up by artist and critic Robert L. Douglas [41] to describe the distinctive African-American formal aesthetics in visual and musical composition that are often dismissed as “too much” under the Western art tradition.

The *Voyager* system embraces the concept of multidominance not only through the formal properties and styles of music it produces — the system is composed of sixty four asynchronously operating voices that all generate music in real time — but also through de-instrumentalizing the computer in the *dialogue* between the human and machine jazz improviser. Its author describes the system as an independent “player” whose performances embody meaning and announce emotional and mental intention.

In contrast to Haraway, Lewis has leveraged the instrumental role of computer in the information-retrieval-based interaction paradigm as a metaphor of underprivileged social classes and cultural forms under the similar rigidity of existing hierarchical structure. In both cases, the role of the computer is both the content and the medium, through which social and political issues are magnified.

2.1.5 Narcissus' Mirror

As we are busy figuring out the role of computer and how our society can benefit from it, computers are changing who we are both literally and conceptually. First, this change happens at a biological level. In the discourse of cyborgs/posthuman, claims have been made that once we implant computers into our body, humans will morph into cyborgs, or that one day we will be able to directly download our conscious thoughts into a computer[151]. A more moderate position is that even the regular use of computers, and any other tools at large, affects the boundaries of our bodies [25]. In other words, we have always been cyborgs throughout the history of tools and technology.

Admitting that “the prospect of becoming posthuman both evokes terror and excites pleasure,” N. Katherine Hayles’ analysis [87] reveals a prominent cultural change underlying the cyborgs/posthuman dream — a shift from the paradox of presence/absence (embodiment) to the one of pattern/randomness (virtuality). The posthuman condition is built on the cultural perception of *virtuality* that material objects are interpenetrated by disembodied information patterns, as opposed to the humanly embodied experience. To Hayles, part of the pleasure of becoming posthuman lies in “getting out of some of the old boxes and opening up new ways of thinking about what being human means.” The dialect of pattern/randomness offers a system that is not front-loaded with meaning and illustrates the emergent power, in which subjectivity truly lies. “The posthuman need not be recuperated back into liberal humanism, nor need it be construed as antihuman”(pp.286), Hayles concludes at the end of her book, as long as we don’t forget that we are first and foremost embodied beings.

What is equally revealing is the metaphor that computer is the mirror through which we understand our own identity. These machines provide the modern version of Narcissus with a surface from which he falls in love again with his reflection. Based

on Jean Piaget’s theory [172] that children develop the concept of life by making finer and finer distinctions about the kind of activity that is evidence of life, Sherry Turkle [203] conducted a series of empirical studies on children’s interactions with computers. The experiments have shown that the computer influences the forging fundamental categories in children’s sense-making of the world because it clearly stands in between the living and the non-living. It sometimes “cheats,” and yet it runs on a battery. It is “programmed,” but so are people. Turkle observes that one position that children often come to is that computers are like people in their *psychology* but not like people in their origins. It is of particular importance to intentional systems because the term “psychology” here is almost a synonym of intentionality.

A particularly powerful moment she described is when one of the children confronted an unplanned bug in a computer toy — the machine simply would not shut down even if the child tried to. Turkle commented that it was a scary moment because it violated what we assume about machines — we can turn them off at will. So at that movement of discovery, the toy became alive, and for our purpose, intentional. The malfunction made its unpredictable from the design stance established by the basic HCI interface design principles and forced people to move on to the intentional stance.

The new categorization challenge is not only limited to children; the arrival of computers destabilized the long-standing dichotomy between the worlds of physics and psychology, set up by Western science. What Turkle sees as more valuable in Artificial Intelligence is that it offers “a new way of thinking. The new way of knowing asks that you think about everything, especially all aspects of the mind, in computational terms, in terms of program and information processing” (pp.246).

This, of course, is not the first time in history that our identity as the human race has evolved in response to confrontation with new discovery and technology. When Copernicus ousted Earth from its illusionary central place in the universe, he

called into question the human identity as God’s privileged servant; when Darwin claimed that humans are descended from animals, he destroyed the image of human as the moral creature with souls transcending all animals. Once more, in the presence of computers, human beings have been deprived of its privileged place as the only entity that *thinks*. In contrast to the Weizenbaum-styled rejection, Turkle embraces computers for they give us another chance to “see ourselves differently as we catch sign of our images in the mirror of the machine” [204, pp.9].

Heidegger’s work on technology may help to summarize the various perspectives that we have gone through so far. In his article “The Question Concerning Technology” [89], the “essence” of technology is to “emframe,” that is, it “reveals” by “concealing.” Similar to many other technologies developed throughout human history, computers once again reframe our understanding of human existence. AI and computers certainly reveal who we are by providing us with a new mirror through which we can see ourselves from a different angle. At the same time, we should not forget that they conceal Being — the truth about human existence prior to logical reflection — which cannot be simulated by computers.

Although a large segment of this history is represented by the narratives of the social changes that computers have brought, it does not imply the subscription to a McLuhanian technological determinism view. In his in-depth study of television, Raymond Williams [214] denounced both technological determinism and the symptomatic technology (called social determinism by Winner) views. He consequently proposed a different interpretation:

[It] would differ from technological determinism in that it would restore intention to the process of research and development. The technology would be seen, that is to say, as being looked for and developed with certain purposes and practices already in mind. At the same time the interpretation would differ from symptomatic technology in that these

purposes and practices would be seen as direct: as known social needs, purpose and practices to which the technology is not marginal but central.

[215, pp.14]

So far, I have offered a condensed overview of major perspectives of computing in the last few decades in order to situate my own practice. This study and our expressive production work depicted in Chapter 6 represent an integrated perspective to call attention to social and cultural issues through computers, especially AI technology, as an expressive medium. It is with Williams’ reminder that we continue to the next section, which lays the groundwork for understanding such “intention to the process of research and development” of intentional systems.

2.2 System Intentionality

Intentionality is a topic of longstanding concern in philosophy. It is commonly described as “aboutness” [2, 37, 217] or defined as “that property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world” [189]. Reintroduced by Franz Brentano [18] in the late nineteenth century and taken up by Husserl [104], the concept is understood as the linkage between the “inexistence” of human mental phenomena and the material establishments and states in the world. Intentional mental states, which include beliefs and desires, are not free-floating thoughts, but are always *about* or *directed at* something. In other words, we do not just have beliefs and desires in their abstract forms. Instead, they are *always* about certain states (e.g., I believe that it is going to rain tomorrow) or directed at certain objects (e.g., his desire for a new sports car).

In the mid-nineteen forties, psychologists Fritz Heider and Mary-Ann Simmel [90] conducted an experiment in which they showed participants clips of animated films with simple geometric figures (see Figure 9), animated with different pacing, movement, location and other factors. Despite the minimalistic representation, none of the

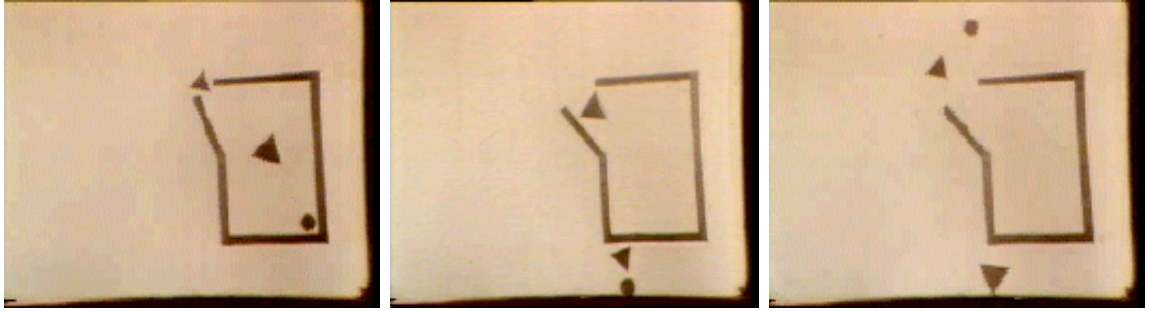


Figure 9: Screenshots of Heider and Simmel’s 1944 Film Experiments

participants saw these triangles and dots simply as mathematical shapes. Similar to the users of Braitenberg Vehicles (Section 1.2), they almost unequivocally reported seeing various types of human-like intentionality in these shapes, such as beliefs, desires and other intentional mental states. Some even developed elaborate narratives, such as that two of the shapes were in love when the third one tried to steal one of them, and that the two lovers eventually defeated their enemy. This important experiment suggests that attributing intentionality is a universal human phenomenon. The phenomenon equips us with the crucial capability of hypothesizing causal relations between various events, characters, and objects, especially when the real cause is beyond our reach.

Despite the universality of the intentional readings developed by Heider and Simmel’s participants, the attribution of intentionality to artifacts is too radical, or even unethical, for many to accept. The rest of this section is structured around the issue of system intentionality. I first present two opposing positions of intentionality raised in the philosophical debates of AI in Section 2.2.1 and 2.2.2 to suggest a broader understanding of intentionality. Next, theoretical developments in several different areas are discussed in terms of how intentionality of artifacts are acknowledged and treated. My goal here is to illustrate that a broader definition of intentionality that does not exclude artifacts is possible and useful to understand intentional systems as they become increasingly pervasive in our culture.

2.2.1 The Chinese Room Argument

The notion of intentionality pertaining to mental states is of particular importance to AI because it directly speaks to the field’s defining goal of building artificial minds. Even though AI has scaled down its ultimate goal of building the thinking mind, recent years have witnessed signs of a revival of this dream. Notably, the Second Conference on Artificial *General* Intelligence was held in Virginia in 2009. Departing from the prevailing pragmatic views of AI described in the last section, this conference intends to encourage research that “focuses on the original and ultimate goal of AI — to create intelligence as a whole.”⁴ This new trend of revival makes the history and lessons from classic AI even more relevant to the present context.

Core debates about AI and intentionality typically center on the question of whether computer programs can exhibit the phenomenon of intentionality in similar ways as humans. A typical attack on the success of early AI takes the flavor of Lady Lovelace’s argument: computers may *appear* to think, but their operations are unintentional execution of instructions and therefore the computers are not and cannot be intelligent in the way humans are. In other words, intentionality is seen as a fundamental prerequisite of human intelligence. At the heart of the debate is philosopher John Searle’s *Chinese Room argument* [188], a controversial thought experiment described below.

Suppose a non-Chinese speaker inside a room is given a large batch of written Chinese symbols and a set of formal rules in his own language to manipulate them. Whenever he receives a message (input) written in Chinese, the imaginary person produces another set of Chinese symbols (output) precisely based on the rules as his reply. This person gets so good at his task that from the vantage point of someone outside the room his reply is indistinguishable from ones produced by native Chinese

⁴The official conference website can be found at <http://www.agi-09.org/>.

speakers. However, this appearance does not change the fact that the person does not speak a single word of Chinese. According to Searle, this action of symbol manipulation is what lies at the core of computational processes. Since the human in the room clearly does not understand Chinese, neither does the computer.

The Chinese Room argument is a direct counter-argument to what Searle calls the *Strong AI hypothesis*, that is, a view that computers and programs can possess intentionality, intelligence, and other mental capabilities in similar ways as humans. The basis of Searle's argument is that a computer's internal states and processes are purely syntactic, as sketched out in Newell and Simon's Physical Symbol Systems Hypothesis (Section 2.1.1). Since the symbols being manipulated are meaningless to the machine under Newell and Simon's framework, the computer therefore is not intentional or, consequently, intelligent.

This argument embodies the long tradition of philosophy of mind wherein intentionality is seen as an intrinsic property of the subject under examination. To Searle, the distinction between the Strong and Weak AI hypotheses represents the difference between intrinsic or original intentionality and intentionality derived through language. He argues that the intentionality of mental states "is not derived from some more prior forms of intentionality, but is intrinsic to the states themselves." By contrast, derived intentionality is "derived from the intentionality of the mind" of the one who interprets them [189, pp.vii]. In the case of digital artifacts, AI systems may hold derived intentionality through being *seen* as playing chess intelligently or understanding natural languages (i.e., the weak AI hypothesis), but they do not have intrinsic intentionality because the symbolic objects and syntactical rules that computers operate on cannot represent beliefs, desires, and other intentional mental states.

Searle's distinction between intrinsic and derived intentionality was certainly an important contribution at the time when the popular culture and government agencies

fell in love with the hyped concept of Strong AI in the context of the Cold War. Nevertheless, it merely rephrases the conventional view that humans are the only privileged form of existence above everything else. More importantly, Searle studies the phenomenon of intentionality as an ontological property of computers in isolation without taking into consideration how people make use of them or the computers' impact on society. As the Strong AI hypothesis fades away from the discourse, Searle's argument is also losing its relevance to intentional systems.

2.2.2 The Intentional Stance

A wide variety of responses have been generated since Searle published his Chinese Room argument. Many of them have been documented in David Cole's article [26]. Daniel Dennett, among others, raises the fundamental question about the existence of intrinsic intentionality, even in human beings. He argues that "we [humans] are artifacts ... designed over the eons as survival machines for genes.... So our intentionality is derived from the intentionality of our 'selfish' genes"[38, pp.298]! For Dennett, all intentionality is derived in the first place. Far from being secondary in Searle's theory, the derived intentionality according to Dennett serves utilitarian functions critical to human survival. The action of attributing intentionality allows us to make sense of our surroundings, an argument not unconnected from the results of Heider and Simmel's film experiment that we see in the beginning of Section 2.2.

Echoing Dennett's theory, Allen Newell's theory of knowledge and its representation [158] stresses that knowledge, a key component to AI, is fundamentally interpretive and derived. Rejecting the possibility that knowledge can be characterized in any (intrinsic) physical *structure*, Newell's theory bases entirely on the *functioning* of knowledge. In a situation where knowledge is used, according to Newell, the notion that an agent has knowledge requires an observer to ascribe it so.

In the previous chapter, I introduced Dennett's theory of the intentional stance,



Figure 10: Roomba Vacuum Cleaner with “Spotty Leopard” Costume

that is, the strategies we undertake to predict a system’s behavior by treating them as if they have beliefs and desires. Recent studies in human computer interaction seem to support this theory. In an empirical study conducted by Sung et. al. [201] on the usage of Roomba (Figure 10), an owner describes his two Roombas, called “discovery” and “scheduler,” in the following way:

And I know definitely they have a same firmware or a similar firmware so the difference should not be much but ah, for example, my discovery, he’s more crazy. He runs into things and sometimes and goes into different places he should not be going to. And the scheduler he’s more like refined. He knows what he’s doing.

Of interest here is that even though the owner is conscious of the technical implausibility of his narrative, he still cannot help reading and describing his robotic vacuum cleaners as intentional, that is, as possessing their own beliefs and desires. A further illustration of the pervasiveness of the intentional stance is Sherry Turkle’s field work on how children interact with computational toys as discussed in Section 2.1.5.

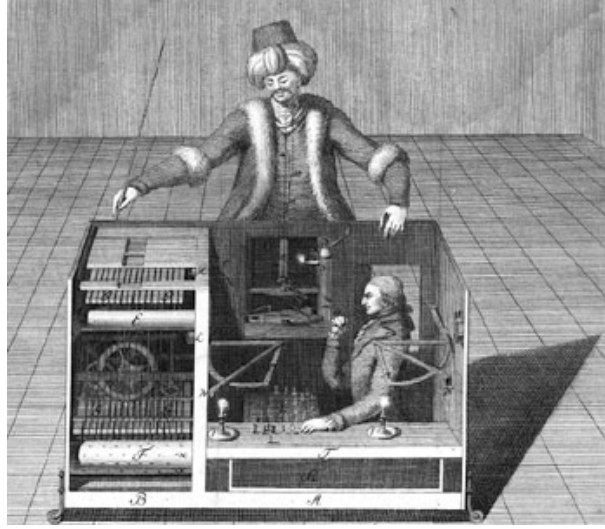


Figure 11: Mechanical Turkish Chess Player by Baron Wolfgang von Kempelen in 1769

2.2.3 The Ghost outside the Machine: Social Perspectives

Dennett’s rejection of the distinction between intrinsic and derived intentionality blurs the fundamental boundary that separates humans from artifacts. The discussion of intentional stance allows us to escape from the ontological debate of intentionality narrowly defined as the sacred property of humans. It also opens up new ways of thinking about artifacts and their role in social settings, which include both humans and artifacts. As radical as it may seem to many people, this view resonates with similar developments in a wide range of other research areas that challenge the traditional dividing line between the intentional subject and the passive object.

It is important to sketch some of these developments in order to contextualize the position of intentionality that this dissertation takes. Although the theoretical and practical explorations of artifacts and intentionality can perhaps be traced back to ancient *automata* (e.g., Figure 11), which literally means “acting of one’s own will” in Greek, this subsection mainly confines its scope to contemporary theories, in particular, activity theory, actor-network theory, and distributed cognition developed within the last century.

As described in the beginning of the chapter, the view that intentionality differentiates us from artifacts manifests itself throughout the long history of Western philosophy, theology and science. Searle’s insistence on intrinsic intentionality clearly embodies this tradition. Since the turn of the twentieth century, however, an increasing number of theorists have questioned this assumption and drawn artifacts closer to the core of humanity than ever.

2.2.3.1 *Activity Theory*

In the nineteen twenties, psychologists in the former Soviet Unions started a new practice with activity as its central unit of analysis. To overcome the drawbacks of psychoanalysis and behaviorism paradigms, activity theorists approach consciousness and intentionality through the analysis of everyday activities that are “firmly and inextricably embedded in the social matrix” of people and artifacts, a broad concept including physical tools or sign systems [155]. Under this framework, artifacts play an active role in *mediating* an individual’s consciousness and actions. For example,

a person may have the object of obtaining food, but to do so he must carry out actions not immediately directed at obtaining food⁵ [He may] make a hunting weapon. Does he subsequently use the weapon he made, or does he pass it on to someone else and receive a portion of the total catch? In both cases, that which energizes his activity and that to which his action is directed do not coincide. [156] (Originally in [127, pp.37])

The lack of coincidence between a person’s conscious actions and intentionality reveals the central mediation role of artifacts: how instruments mediate between an actor and the object of doing; how the object negotiates with the limitation of the

⁵In the terminology of activity theory, an “object” refers to the “objective” of an individual.

instrument; and how artifacts are created and transformed with the development of activities and therefore carry with them particular culture and history [115].

While activity theory portrays artifacts as “integral and inseparable components of human functioning” [116], the framework is still constructed around human consciousness and objectives. This “asymmetry between people and things” [156] is further questioned and eliminated by the other two recent developments in actor-network theory and distributed cognition theories.

2.2.3.2 Actor-Network Theory

Actor-network theory (ANT) was founded from science and technology studies during the early 1980s by Bruno Latour [122, 124] and Michel Callon [23]. Traditionally, the study of human actors is the territory of sociological inquiries. Non-human actors, on the other hand, are often greeted by the attitude to “leave the nonhuman to the care of technologists or to study the impact of black-boxed techniques upon the evolution of social groups” [121]. In contrast, actor-network theory attempts to understand social interactions through the lens of both human and non-human factors.

Even for many liberal scholars, ANT’s insistence on treating human and non-human actors equally in its analysis is an “absurd” move that separates it radically from conventional sociology and many other traditions. A new construct called *theactant* was created deliberately to downplay the distinction between human and non-human actors in the complex networks of associations and translations. Before diving into ANT’s implications for AI and intentional systems, it is worthwhile to contextualize this bold gesture.

In a 1988 article “Mixing Humans and Nonhumans Together” [121], Latour elucidates the foundation of his position. Starting from an extremely simple technological invention of the door, Latour reveals an increasingly complex network in which functions and skills are delegated in an intertwined manner between human actants (e.g.,

a gatekeeper) and non-human actants (e.g., a hinge and a hydraulic door-closer). The traditional sociological approach falls short here in providing a coherent explanation in this simple example, for the field's "absence of a convenient vocabulary hinders its practitioners from mov[ing] freely from studying associations of human to associations of nonhumans." Latour points out cogently,

It is because humans, nonhumans, and even angels are never sufficient in themselves and because there is no one direction going from one type of delegation to the other, that it is so useless to impose a priori divisions between which skills are human and which ones are not human, which characters are personified and which remain abstract, which delegation is forbidden and which is permissible, which type of delegation is stronger or more durable than the other. In place of these many cumbersome distinctions why not take up a few simple descriptive tools?

In the same article, Latour relates his concept of "the distribution of skills between humans and nonhumans" to the low-hanging overpasses on Long Island, a famous example in Winner's article on how artifacts embody forms of power and authority[215]. These unusually low overpasses were designed by Robert Moses, a major builder of public infrastructure in New York between 1920s and 1970s. As part of his political vision of the city, Moses constructed these overpasses "according to specifications that would discourage the presence of buses on his parkways." This means that the lower-class citizens who do not have access to automobiles will not have the means to commute on these parkways and hence lose their access to the connected neighborhoods. Automobile-owning whites of "upper" and "comfortable middle" classes, as Moses calls them, on the other hand, would be free to use the parkways for recreation and commuting. In this particular example, actor-network theorists would argue that the technological objects of overpasses serve similar social functions as human police hired to reinforce segregation laws as well as the immaterial ideology of social and

class inequality.

2.2.3.3 *Distributed Cognition Theories*

Distributed cognition theories represent recent perspectives in cognitive science that expands the field’s traditional emphasis on the mental processes occurring inside an individual’s brain. Similar to the above two theories, distributed cognition pays attention to the close interaction between human and artifacts in a social-cultural system. As defined by its founding members, distributed cognition perspectives address “the representation of knowledge both inside the heads of individuals and in the world ...; the propagation of knowledge between different individuals and artifacts ...; and the transformations which external structures undergo when operated on by individuals and artifacts” [56].

In parallel with actor-network theory, distributed cognition approaches also adopt a more symmetric relationship between human and non-human by assigning them theoretical equivalence as “agents” in their analyses. More importantly, both forms of agents are necessary components of distributed cognition approach’s unit of analysis — the cognitive system. The theory studies the cognition of a system in terms of its functioning, emphasizing the organization, coordination, and operation among its components.

Hutchins [106, 156] provides an example of a cognitive system of an airplane cockpit, where the system’s goal is the “successful completion of a flight.” In order to understand the “cockpit system,” we have to consider the *unity* of pilots and instruments because “the properties of individual agents alone, no matter how detailed the knowledge of the properties of those individuals might be” is inadequate. The human and non-human components of the cockpit form a single cognitive system and coordinate with each other to enact the goal.

In a more detailed study, Hutchins [105] analyzes a cognitive system of shipboard

navigation on Navy vessels. He contrasts Western navigation, which relies on artifacts to aid the procedure of locating the navigator's position on a chart, with Micronesian navigation, which has used the reference of star sightings, speed judgments, and other contextual information. Hutchins argues that although the computational problem facing both Western and Micronesian steersmen is the same, the representations and processes to carry out that task (algorithmic level) and the physical instantiation (implementational level) are different in the two cases. Western navigation is algorithmically complex with details and calculations. However, much of these tasks are pre-calculated and embedded in instruments and procedures such as charts and devices. The cognitive system of navigation, including the navigation team together with artifacts and procedures, organizes itself towards settings that are robust and cognitively economical to navigators.

In summary, this section surveyed multiple contemporary theoretical investigations of intentionality in the context of artifacts. Starting with the critiques of AI in philosophy of mind, we first encountered Searle's Chinese Room argument that machines may have derived intentionality, which should not be confused with human intrinsic intentionality that computer symbol manipulation processes can never achieve. The distinction between derived and intrinsic intentionality has been challenged, among others, by Dennett who questions the existence of the so-called intrinsic intentionality, even in human beings. Instead, he has proposed that *all* intentionality is derived, as a result of the human survival instinct to predict the behaviors of other elements in the environment, both the living and not, especially when little information is known about them. The ontological debates about intentionality in the traditional sense were useful in their particular context when the field of AI was inflated with unrealistic ambitions that baffled people's perception on humanity at the time. However, as stated in the previous section, the ontological question of "Can machines think?" has lost its importance in society. It is more constructive, in my

view, to adopt broader understandings of intentionality that allow us to decipher the roles of technological artifacts, as they become increasingly pervasive in our society. We hence briefly visited three major theoretical developments in the past century — activity theory, actor-network theory and the distributed approach — that form the theoretical foundation of thinking through the concept of system intentionality later in Chapter 3.

2.3 *Reading Systems*

The task of analyzing intentional systems in terms of how system intentionality is formed requires close readings of these systems. This section presents my theoretical framework built on the existing theories of “reading.” First, 2.3.1 offers a succinct account of the hermeneutic tradition of interpretation as the foundation of my main method. One particular school of hermeneutics, as we shall see, is also of particular importance to AI and intentional systems because of its philosophical concerns on human existence and intelligence. Next, an emerging research area that I find to be surprisingly closely related to hermeneutics called “software studies” is described in Section 2.3.2. Section 2.3.3 then introduces cognitive semantics as an alternative to read intentional systems. Finally, in Section 2.3.4 examines the existing analytical tools from the field of AI and human computer interaction, two areas closely related to intentional systems, and points out the drawbacks for each analysis schema to fully understand system intentionality.

2.3.1 Hermeneutics

The venerable field of hermeneutics covers the theory and art of interpretation. Preoccupied with questions regarding interpretation, understanding, and point of view, the field has gone through several significant shifts. *Classical methodological hermeneutics* is devoted to methods of studying texts, often sacred and mythical ones. In contrast,

hermeneutics since the beginning of the twentieth century has been focusing on philosophical inquiries; *philosophical hermeneutics* and its various extensions are primarily concerned with the philosophical foundation of meaning and understanding.

The field of hermeneutics bears a long tradition that stretches back to ancient Greek philosophy. A comprehensive overview of the field is well beyond the scope of this section. More thorough studies can be found in [164, 176, 208, 174]. Instead of offering a thorough and authoritative overview of hermeneutics, the goal here to provide necessary background to understand the major critiques of classic AI from philosophical hermeneutics-based perspectives. More importantly, this section prepares the reader for the method of reading intentional systems in Chapter 3, which is informed by classical hermeneutics.

2.3.1.1 Classical Methodological Hermeneutics

As a major branch of biblical study during the middle age and Renaissance, classical methodological hermeneutics (classical hermeneutics in shorthand) primarily concerns with developing systematic and interpretive methodologies to retrieve the author's intended meaning from the text. During the middle age, interpreters of the biblical text based their understandings on empathetic projection instead of reading the sacred text literally. This approach was further developed and systemized by Schleiermacher and Dilthey into a coherent methodology for text interpretation in general. Schleiermacher [164, 186] combined grammatical analysis from classical philology with the study of themes from Biblical exegesis and developed a general framework of text interpretation.

Expanding Schleiermacher's theory, Dilthey [39] claimed that interpreting texts, verbal utterance, actions, and art requires the action of recognizing the intention behind these meaningful expressions. This action is more like interpreting a poem

than conducting physics or chemistry experiments. This viewpoint is well summarized by modern day scholar Henry Virkler's description on the duality of classical hermeneutics:

[Hermeneutics] is considered an art because communication is flexible, and therefore a mechanical and rigid application of rules will sometimes distort the true meaning of a communication. To be a good interpreter one must learn the rules of hermeneutics as well as the art of applying those rules. [208, pp.16]

Dilthey also stressed the importance that lived experience (Erlebnis) and understanding (Verstehen) have in the methodology of the humanities, in contrast to the natural sciences. Erlebnis is connected with the process of self-understanding, whereas Verstehen relates to the understanding of others. Dilthey argues that Erlebnis alone does not provide self-understanding. The latter is only obtained to the extent that the self relates to itself as it relates to others.

One of the underlying assumptions of classical hermeneutics is that the author's intended meaning is the only valid interpretation. Therefore, a hermeneutical interpreter's goal is to overcome the barrier of the passage of time, the evolution of language, and the change of context, and to successfully retrieve the true meaning intended by the original author(s) of the text. This view is still shared by the modern followers of classic hermeneutics. In his major work *Validity in Interpretation* published in 1967, Hirsch expressed his strong objection to relativism:

When critics deliberately banished the original author, they themselves usurped his place [as the determiners of meaning], and this led unerringly to some of our present-day theoretical confusions. Where before there had been but one author [one determiner of meaning], there now arose a multiplicity of them, each carrying as much authority as the next. To

banish the original author as the determiner of meaning was to reject the only compelling normative principle that could lend validity to an interpretation For if the meaning of a text is not the author's, then no interpretation can possibly correspond to the meaning of the text, since the text can have no determinate or determinable meaning. [92, pp.5]

In their persistent pursuit of *the* valid interpretation, hermeneutic interpreters quickly found themselves stuck in a difficult situation — the understanding of text as a whole is dependent on the knowledge of its constituting parts, which in turn is equally relying on the interpreter's comprehension of the entire text. This issue of the *hermeneutic circle* is one of the major challenges and themes of the field.

2.3.1.2 *Philosophical Hermeneutics*

Under the influence of German romanticism and idealism, the field of hermeneutics saw a major shift of focus from being other disciplines' methodological or didactic aid to philosophical inquiries of “the conditions of possibility for symbolic communication as such” [174]. Heidegger and Gadamer's *philosophical hermeneutics* raised the notion that interpretation is not based on knowing, but based on our direct and non-mediated being-in-the-world. The field's new task during this period is well summarized by Ricoeur's definition that “hermeneutics is the theory of the operations of understanding in their relation to the interpretation of texts” [176, pp.43].

Whereas Dilthey's hermeneutics primarily tackles the epistemological problem of communicating with other people, Heidegger's approach links it to the ontological foundation of existence in relation to the world through the concept of *Dasein*. In his influential work *Being and Time* [88], Heidegger defines *Dasein* as direct, non-mediated being in the world, “a being with being.” Its constitution cannot be displayed through the conventional human sciences (*Geisteswissenschaften*), which are “ground[ed] by derivation.” In other words, *Dasein* exists prior to scientific knowing.

Heidegger's philosophical hermeneutics is extended by Gadamer particularly in his *Truth and Method*[67]. Gadamer links the nature of being into Heidegger's notion of *thrownness*, that is, the condition in which human beings are "thrown" into the world independent of their wills and cannot step back and reflect on their actions. Gadamer proclaims:

To acquire an awareness of a situation is, however, always a task of particular difficulty. The very idea of a situation means that we are not standing outside it and hence are unable to have any objective knowledge of it. We are always within the situation and to throw light on it is a task that is never entirely completed. This is also true of the hermeneutic situation, i.e., the situation in which we find ourselves with regard to the tradition that we are trying to understand. (pp.268)

2.3.1.3 *Hermeneutics and AI*

Within the work that has been done with hermeneutics in the context of AI, attention has been so far primarily concentrated on the philosophical hermeneutic tradition. It is not surprising given that Heidegger and Gadamer's theory of human existence and understanding fundamentally contradicts the approaches of classic AI. Dreyfus' 1972 work *What Computers Can't Do: A Critique of Artificial Reason* [42], for example, ardently criticizes the plausibility of AI on hermeneutic grounds. He claims that the foundation of AI is built on biological and psychological assumptions as well as deeper epistemological and ontological assumptions. With these assumptions, AI and cognitive scientists came up with the view that cognition is the manipulation of internal symbols specified by internal rules. Using Heidegger's hermeneutics as his theoretical basis, Dreyfus argues that understanding is fundamentally based on our embodied and highly context-bounded being-in-the-worldness and thrownness, none of which can be formalized into rule-based symbol manipulation. For this reason,

according to Dreyfus, classic AI is doomed to fail.

In the domain of natural language understanding, AI veteran Terry Winograd and philosopher Fernando Flores [217] reached a similar conclusion by combining hermeneutics with biological and technical vantage points.⁶ They take on the Western rationalism understanding of language, thought, and rationality, and show how its influence sometimes blinds us from seeing other alternatives. Based on both Heidegger’s hermeneutics and Maturana’s biological study of vision, the authors remind us that meaning and Being require flexible accommodation of their ever-changing context.

Remarkably, some studies on hermeneutics have been also done inside the AI community. Mallery, Shurwitz and Duffy [133] from MIT’s AI lab, for instance, conducted an investigation on hermeneutics, hoping to find insights to understand “meaning, translation, architectures for natural language understanding, and even ... methods suitable for scientific inquiry in AI.” Although the authors noticed the parallel between the hermeneutic circle and a computational notion of bootstrapping, the article does not offer any concrete suggestions for the possible uses of hermeneutics in AI.⁷ This, in a way, reflects how deeply the scientific and technological fields have invested in the rationalist tradition. As Winograd and Flores have pointed out, anything that diverges from this deep-seated preunderstanding of rationalism risks the possibility of doubt and rejection [217, pp.16]. However, Mallery et. al.’s work is certainly important for the field of AI in that it is a sincere and constructive response to many hermeneutic-based critiques that it has received.

Before we move on, it is worthwhile to mention that Rodney Brooks’s interactionist AI, a more recent paradigm of AI, is seen as bearing resemblance to Heidegger’s theory.

⁶Frustrated by the fundamental issues of AI described in this joint work, Winograd has shifted his focus to HCI.

⁷Bootstrapping is a computational concept that uses lower order components to construct higher order ones, which in turn modify the former.

Compared with the classical AI paradigm, interactionist AI approaches intelligence as an emergent phenomenon resulting from an agent’s “physical interaction with the environment” [20]. Even though Brooks himself denied the influence, one cannot help seeing the close connection between the two fields.

The two major developments of hermeneutics described above offer insights to different aspects of meaning and understanding. The classic hermeneutics focuses on the methodology of communication through textual interpretation, whereas its philosophical counterpart deals with ontological foundation of understanding through being-in-the-worldness. Their distinctions lend themselves to different applications in the context of AI. As we have seen, the philosophical hermeneutics has proven to be a powerful tool to examine the direction and approaches of classic AI. The classical hermeneutics can also be useful for the purpose of understanding these systems through close readings, as illustrated in Chapter 3. The rest of the section introduces two other approaches that will prepare us to read intentional systems. Software studies, to a certain extent, is related to classical hermeneutics in its approach to read software systems as text. Cognitive semantics research, on the other hand, provides a particular way of close reading that is not currently covered in software studies, that is through the discovering of entrenched metaphors and blends in the texts.

2.3.2 Software Studies

The emerging research area of software studies, whose name was coined in Lev Manovich’s book *The Language of New Media*[134], applies the theories and approaches from humanities and social sciences to computer software. It is related to the classical hermeneutics in that scholars of software studies perform close reading of computer software in order to gain insights of its authors as well as the embodied social and cultural contexts in which the code was written. To put it succinctly, computer software is studied as text.

In the website of the recently started “Software Studies Initiative” at University of California, San Diego, its creators Lev Manovich and Noah Wardrip-Fruin put forth their motivations for the initiative:

[W]e think of software as a layer that permeates all areas of contemporary societies. Therefore, if we want to understand contemporary techniques of control, communication, representation, simulation, analysis, decision-making, memory, vision, writing, and interaction, our analysis can’t be complete until we consider this software layer.⁸

The origin of software studies is linked to the debate in late nineteen nineties between software formalism and software culturalism. Commenting on Geoff Cox et. al.’s widely read manifesto proclaiming that “the aesthetic value of code lies in its execution, not simply its written form” [30], Florian Cramer [31] labeled Cox and his supporters as *software formalists* as they “focu[s] on the formal poetics and aesthetics of software code and individual subjectivity expressed in algorithms.” This is to contrast with another approach of those whom Cramer calls the *software culturalists*. As their name suggests, this group of people regard software as, first and foremost, a culturally and politically coded practice and focus their attention on addressing these aspects of software. Earlier influential work in this lineage includes Matthew Fuller’s work on software as culture and his cultural analysis of the Microsoft Word interface [65], Paul N. Edwards’s study on the impact of the closed world ideology during the Cold War political environment on the development of computer technology [45], and Friedrich Kittler’s claim that the concept of software is itself an illusionary construction under the influence of Cartesian Dualism [113]. These early software culturalist works laid the ground for what was later called software studies.

Many new generation software studies scholars are also influenced by the software

⁸Excerpts from the official website of Software Studies Initiative: <http://lab.softwarestudies.com/2007/05/about-software-studies-ucsd.html>

formalists in the sense that both stress the importance of procedural literacy. Instead of blackboxing the underlying execution of software, these scholars see programming and source code as the primary subject of their analysis. In his book *Software Studies: A Lexicon* [66], the first book dedicated to software studies, Fuller collected a wide range of articles covering topics such as algorithms, data structures, computing languages, data visualization and interaction. Although this list of topics may seem quite common for a computer science and software engineering book, Fuller’s book represents a range of vastly interdisciplinary approaches. Compared to the early software culturalists, the book’s contributors, many of whom have extensive background in computing, unseal the blackbox of software and analyze these seemingly technical issues with critical and cultural lenses.

Meanwhile, other scholars remind us to look outside of the box of programming for its own sake. For example, Kirschenbaum’s article [112] titled “Virtuality and VRML: Software Studies After Manovich” traced the downfall of a once-promising virtual reality programming language VRML to its business and managerial causes. He points out that while studying the intangible software, we should not lose sight of its “materialistic” causes. For Kirschenbaum, software is the product of “white papers, engineering specs, marketing reports, conversations and collaborations, intuitive insights and professionalized expertise, venture capital (in other words, money), late nights (in other words, labor), Mountain Dew, and espresso.”⁹

As a young and rising research area, software studies represents a vastly different and sometimes contradictory approach. However, its nucleus of applying humanistic close reading to software systems both at their source code and their developing

⁹An interesting parallel inside the field of AI is Rychtyckyj and Turski’s [182] realization that the failure of most commercial AI expert systems was not due to technical problems, but mostly to the company’s organizational and business issues. This paper was presented last summer at the *What Went Wrong and Why* workshop in the AAAI 2008 conference. It is an important landmark, even though Latour has made the very same argument in Latour [124], because the realization was made *inside* the AI community of practice.

context is useful to the analysis of intentional systems.

2.3.3 Cognitive Semantics

The last corner stone that lays the theoretical framework for reading intentional systems is drawn from cognitive semantics, particularly from its areas of conceptual metaphor theory (a.k.a., metaphor theory) [118] and conceptual blending theory (a.k.a., blending theory) [53, 54]. As part of the broader cognitive linguistics movement, both metaphor theory and blending theory scrutinize the usage of figurative languages including metaphors, metonymies, and blends in linguistic data with the goal of gaining insights of how human cognition functions. These figurative languages are so entrenched in and coherent with our culture that some believe that they provide cognitive structures for our worldviews [55]. With it comes the realization that these entrenched linguistic patterns in discourses are important to recover fundamental cultural and ideological beliefs. In addition, certain components of metaphor and blending theories suggest computational models that can be used to algorithmically generate figurative expressions that have both cultural and aesthetic significance [81]. The following two sections provide brief accounts of both theories and how they fit into my overall framework.

2.3.3.1 *Conceptual Metaphor Theory and Its Application*

Conceptual metaphor theory originated in George Lakoff and Mark Johnson’s 1980 seminal work *Metaphors We Live By* [118]. The authors assert that metaphor is not just a matter of language, but more importantly, their pervasive manifestations reflect the fact “[o]ur ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature” (pp.3). Through various examples, Lakoff and Johnson demonstrate that human understandings are structured coherently with the conceptual categories, or gestalt understandings, we form through our recurrent bodily experiences. These experiences are understood *metaphorically* when we use

a gestalt understanding from one domain to structure experience in another. For instance, the multi-dimensional structures of the concept *LOVE* are partially based on the more concrete concept of *JOURNEY*. In a later article, Lakoff [117] suggests that metaphors are organized in a hierarchical structure, from the most fundamental and cognitive ones, which are grounded in our pre-conceptual embodied experiences, to higher level conceptual and cultural ones.

A metaphor, within the framework of Lakoff and Johnson, involves the mapping between two conceptual structures, a *source domain* and a *target domain*. The source domain is from which we draw metaphorical expressions, and is typically (but not always) composed of a concept that is familiar and understandable directly through human embodied experience (e.g., *JOURNEY*). The target domain, on the other hand, is the conceptual domain that we intend to understand and is often less immediate to our sensory-motor skills and more abstract (e.g., *LOVE*). The metaphorical *mappings* from the source to target domain establish correspondences between the two structures in many literary and everyday expressions, so that people can better grasp the more abstract target domain. For example, the metaphor *LOVE IS A JOURNEY* is manifested through *text expressions* such as “I don’t think this relationship is *going anywhere*,” “Look *how far we’ve come*,” and “We’re *at a crossroad*.”

Metaphors embody worldviews through its *metaphorical systematicity* of highlighting and hiding [118, pp.10]. Different metaphors bring forth correspondent aspects of the target domain. For instance, the metaphor *LOVE IS A PHYSICAL FORCE* describes the concept differently from *LOVE IS MAGIC* or *LOVE IS A PATIENT*. As these metaphors provide insights through highlighting certain aspects of the target, they also hide others. As Lakoff and Johnson point out, “it is important to see that the metaphorical structuring involved here is partial, not total. If it were total, one concept would actually *be* the other, not merely be understood in terms of it” (pp.13).

In addition to its apparent cognitive contribution, metaphor theory has important social and critical implications. It offers a unique lens for critical studies precisely because certain fundamental metaphors are so deeply entrenched in our culture that most of time people are not conscious about their impact. Otto Santa Ana [183], for instance, combined metaphor theory and Foucault’s critical discourse analysis to reveal the hidden political ideology and social order regarding Latino immigrants in U.S. though analyzing metaphors used in mass media during a specific political era. Santa Ana’s work has proven to be particularly inspirational for unpacking how AI systems’ intentionality is constructed through metaphorical projections in the corpus of technical literature. It is hence worthwhile to pause for a moment and give a brief description of his methods as the methods used in Chapter 4 draw on this work.

Santa Ana and his team first gathered their corpus of texts published by the *Los Angeles Times* from 1992 to 1998, a period that coincides with the campaign period of several important propositions, namely Proposition 187, 209, and 227, regarding illegal (Latino) immigrants in the state of California. From these articles, they identified and isolated non-literal linguistic expressions pertaining to the characterization of Latinos and related topics. The data was then compiled into a table for further analysis, regarding every single text expression of a metaphor, which Santa Ana calls metaphor tokens.

Each metaphor token was then analyzed in the context of the complete article through interpretive classification of its source and target domains by more than one person. Santa Ana’s team members then took the painstaking task of iteratively sorting and editing these labeling/coding generated from the previous procedure until the major typological dimensions of the source/target conceptual domain became clear.

The result of the analysis was enlightening. Through the vast amount of data under examination, Santa Ana located a single metaphorical mapping, IMMIGRATION

AS DANGEROUS WATERS, which was dominantly repeated. From this finding, Santa Ana concludes that “while metaphors are at times consciously used for stylistic purposes, these oratorical functions are eclipsed by their conceptualizing and signaling functions” (pp.58-59). Santa Ana’s method plays an important role in this dissertation, for it influences my method of analyzing an AI system, *Copycat*, discussed later in Chapter 4.

2.3.3.2 *Conceptual Blending Theory*

The more recent conceptual blending theory [54], building upon Gilles Fauconnier’s mental spaces theory [51], describes the means by which concepts are integrated. The theory describes how we arrive at new concepts through blending partial and temporary pieces of information. Most importantly, it proposes that conceptual blending processes occur uniformly in pre-conscious everyday thought and in more complex abstract thought such as in literary arts or rhetoric. Though the empirical grounding for such findings is still being developed [69], this turn toward backstage cognition reflects the trend in cognitive science away from formal, logical, and rational thought and toward context-driven, cultural, and embodied thought [52]. Conceptual blending theory is often seen as a rival with the closely related conceptual metaphor theory. An in-depth comparison between the two approaches can be found in Grady’s paper [73] and will not be repeated here.

Conceptual blending theory serves two purposes for this dissertation. First, it is used as an analytical tool, in addition to metaphor theory, to decode the narration of intentionality in intentional systems. It provides constitutive principles (comprising an idealized model and process) and governing principles (constraints determining which conceptual blends are more “optimal” than others for everyday thought) for how blends are constructed [54]. Second, its computational models, in particular Harrell’s Alloy algorithm [81], provide a generative source to the *Memory, Reverie*

Machine (*MRM*) project.

Conceptual blending is generally regarded as a basic and ubiquitous human cognitive operation that manifests itself from the simple natural language phrases [54] to the more complex mental activity of what Mark Turner calls double-scope stories:

Consider the as yet unexplained human ability to conjure up mental stories that run counter to the story we actually inhabit.... Or suppose you are actually boarding the plane to fly from San Francisco to Washington, D. C. You must be paying attention to the way that travel story goes, or you would not find your seat, stow your bag, and turn off your personal electronic devices. But all the while, you are thinking of surfing Windansea beach, and in that story, there is no San Francisco, no plane, no seat, no bag, no personal electronic devices, no sitting down, and nobody anywhere near you. Just you, the board, and the waves. [206]

Such mental activities, for our purposes, can also be seen as daydreams [81] — the subject of *MRM*. In this computational narrative project, daydreams, memories, and affective dispositions are generated differently for each output through the construction of blends. In particular, Fox Harrell’s approach to computational narrative [80, 83] and his technical work such as his *GRIOT* system and its constitutive Alloy conceptual blending algorithm [81, 71] (Section 2.4.2) are important components of the *MRM*’s technical foundation.

2.3.4 AI and HCI Analyses

For the last three decades, composer and scientist David Cope has been using his *Experiments in Musical Intelligence* — a computer program he initially developed in 1981 — to generate musical compositions in the styles of well-known classical composers as well as to help him compose over 6,000 pieces. He has subsequently released five albums with the titles of *Bach by Design*, *Classical Music Composed by*

Computer, Virtual Mozart, Virtual Bach, and Virtual Rachmaninoff. To my untrained ear, it is hard to miss the similarity between the computer generated music and that written by the classical composers after whom those albums are named. Even some professional musicians have mistaken the music the *Experiments in Musical Intelligence* (*EMI*, pronounced “Emmy”) generated as an unknown piece by Mozart.

At the surface level, *EMI* and George Lewis’ *Voyager* (Chapter 1) share many similarities; they both algorithmically generate music and can both be seen as intentional. As described in Section 2.2, system intentionality refers to the phenomenon of inanimate artifacts such as computational systems behaving as if their actions were *about* something in the world rather than the mere execution of algorithmic rules. These systems are *seen* as exhibiting complex behaviors commonly regarded as the territory of intentional human phenomena (e.g., planning, learning, and creating), which are derived from mental states such as belief and desire. This chapter attempts to locate the formation of system intentionality.

A more detailed look uncovers the different sets of values and practices that the two systems speak to. Unlike Lewis, Cope does not claim that his program should be seen as an independent player or an advocate for oppressed non-Western cultures. On the contrary, the intentionality of *EMI* is proportional to how well it replicates the icons of elite Western culture. The popular stories of how *EMI* “tricks” the listeners is revealing. They illustrate Cope’s close linkage to the field of AI by naming his project “musical intelligence” and framing the goal of his program as a version of the “musical Turing test.” George Lewis, on the other hand, does not align himself with AI or scientific practice. It is true that Lewis [165] admits his ambition to “get the computer [program *Voyager*] to sound good more or less on its own, so that someone listening to it says, ‘*Who* is that playing?’ ” (emphasis added). However, Lewis and Cope deploy very different strategies in the stories they create about their respective systems: Lewis narrates his system as an independent player, not as an imitator of

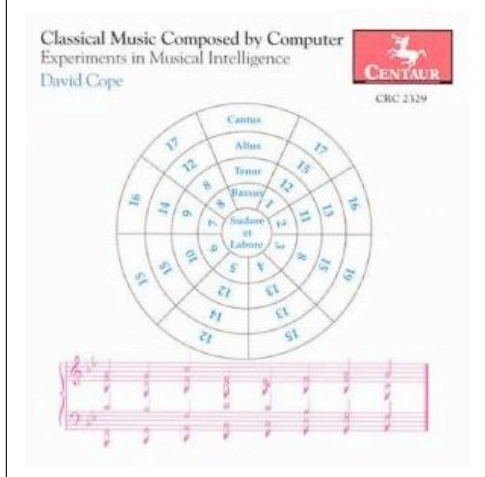


Figure 12: CD Cover Art of Cope’s 1997 Album *Classical Music Composed by Computer: Experiments in Musical Intelligence*

any human beings. Cope, on the other hand, takes advantage of the aura of a classical composer in order to construct a narrative that *EMI* is as good as one of the greatest human composers.

The different approaches that authors take greatly affect how their systems are perceived. As we see from the example of *EMI* and *Voyager*, the authors’ deeper cultural and political stances, the way that they frame the problem, and how they talk about it publicly on the website or through interviews, all shape how we see and make sense of their systems. As we will see in the rest of this section, however, none of the above is taken seriously into account either in AI or HCI analysis, the two most relevant existing analysis schemas for intentional systems.

2.3.4.1 An AI Approach

Preoccupied with the concern for performance and generalizability, the AI approach concentrates on analyzing the formal properties of the system itself. Based on the established conventions of the main areas of AI (e.g., knowledge representation, planning, and machine learning), this section describes the core metrics that AI practitioners use in their analysis. Specifically, these methods are concerned with knowledge representation, algorithmic complexity, domain knowledge, and performance (Table

Table 3: AI Approach for Analyzing Systems

Aspects	Evaluation Criteria
Knowledge Representation	expressivity vs tractability
Domain Knowledge	amount of domain knowledge the system has
Algorithmic Complexity	CPU time, and memory space required
Performance	how well its actual execution achieves the tasks
Architecture	how well a specific cognitive process is carried out

3).

Consider an AI practitioner who encounters an intentional system, such as *Voyager* or *EMI*, and wants to analyze it by employing the AI approach. From a knowledge representation (KR) point of view, she will primarily consider the “expressivity” (sometimes referred to as “expressiveness”) and “reasoning requirements” of how knowledge is represented in the system. A well-known challenge in the KR community is the “tradeoff between expressiveness and tractability,” as described in the textbook by Levesque and Brachman [16]. Discussing the impact of knowledge representation languages on AI algorithms, the authors wrote:

Basically, the problem is that it can be more difficult to reason correctly with one representational language than with another and, moreover, that this difficulty increases dramatically as the expressive power of the language increases. This leads to a tradeoff between the expressiveness of a representational language and its computational tractability. Here we show that this tradeoff can be seen to underlie the differences among a number of existing representational formalisms, in addition to motivating many of the current research issues in knowledge representation. [128]

Notice that the term “expressive” in KR has a very different meaning from its usage in the humanities and the arts. Expressivity in KR refers to the *range* of different concepts that can be represented with a particular knowledge representation. For instance, a simple representational formalism may be just a sequence of musical

notes. It is useful to formalize certain musical pieces, but this representation is confined to considering only one note at a time. A more expressive representation may incorporate multiple parallel sequences of notes. While the second representation can represent a larger variety of musical experiences, the algorithms involved will have to be correspondingly more complicated and therefore more expensive.¹⁰ Thus, in the practice of AI, the selection of knowledge representation is a tradeoff between its expressivity and the consequent computational cost.

Next, the AI practitioner may be curious about the extent of domain knowledge a system requires. In general, the amount of knowledge is often inversely proportional to the complexity of the algorithm needed to solve a problem. In a planning system, for example, domain knowledge can narrow the choice of possible plans and thus makes the search easier [6]. Similarly, domain knowledge in a machine learning system constrains the set of valid hypotheses, which eases the learning task [24]. An extreme scenario in the domain of computer music is to equip the system with a database full of already composed songs and specify which one to play for any given input. High as the quality of the system “produced” music may be, the system is not interesting by AI standards because the complexity of the algorithm involved is trivial. On similar grounds, influential AI researcher Drew McDermott recently rejected the idea of passing the Turing test with an exhaustive response table [86] as valid AI research.

11

As we have seen already, in addition to the expressivity of a system’s knowledge representation, another important measurement of the AI system is algorithmic complexity. One part of algorithmic measurement is the ratio of all acceptable solutions to the entire output space. To put it simply, if all possible solutions in the output

¹⁰In this computer science terminology, “expensiveness,” sometimes also called “computational cost,” or “cost,” refers to the number of CPU cycles (time) and amount of memory needed to execute a certain algorithm.

¹¹This unpublished article can be found in McDermott’s website: <http://cs-www.cs.yale.edu/homes/dvm/papers/turing-table.html>

space are equally good, the algorithm does not have to do much more than pick one at random. On the contrary, if the algorithm has to find the only valid solution out of tens of thousands of possible ones, a situation commonly referred to in the AI community as a “needle in the haystack,” its complexity needs to be remarkably higher than in the previous case.

Although high algorithmic complexity is a challenging AI research problem by itself, it is also associated with the high cost and low efficiency of the algorithm. This aspect is especially important for algorithms that require a lot of computer memory to function. In standard computer science analysis, computational complexity is formally defined as the number of cycles that a Turing machine executing the algorithm would take to solve a problem with respect to the size of input [29], commonly notated as the big “O.” In certain AI areas, such as planning, reducing computational complexity is one of the primary concerns. For example, in the biannual *International Planning Competition*, held in conjunction with the International Conference on Automated Planning and Scheduling (ICAPS), the evaluation criteria for success are “minimize total time” (computational complexity) and “minimize total cost” (smallest number of actions in the proposed plan).¹²

Besides the above theoretical metrics, all AI programs have to run on computers, process actual data, and be analyzed in terms of their actual performance — how well they accomplish a pre-determined task according to relevant measurements. In AI research, typical generic measures of performance include “classification accuracy” in machine learning and, as previously mentioned, the computational complexity measures in the case of planning. Even though these objective metrics continue to dominate the field of AI, there have been critiques [120, 43] from the community pointing out that a single numeric measure of accuracy does not adequately reflect the scientific interestingness of an algorithm. When an AI practitioner tries to measure the

¹²Official website of the competition can be found at <http://ipc.informatik.uni-freiburg.de/>

performance of systems in the expressive domains, such as *Voyager* and *EMI*, bigger issues surface, namely, how to objectively ascertain such subjective measurements as quality of music?

Admittedly, there have been a number of attempts to formalize and objectify subjective measures such as creativity and expressiveness towards the goal of applying AI analyzing methodology to expressive domains including music, narrative, and visual arts. Pérez y Pérez and Sharples [169], for instance, propose the construct of “computerised creativity” (c-creativity) in the domain of story generation based on Boden’s work [13] on creativity. In her book *The Creative Mind*, Boden defines psychological creativity (p-creativity) as “ideas ... that are fundamentally novel with respect to *the individual mind* which had the idea” (pp.32) and differentiates it from historical creativity (h-creativity), which concerns the entire human history. Applying the concept of p-creativity, Pérez y Pérez and Sharples propose to measure c-creativity in terms of how much a system “generates knowledge that does not explicitly exist in the original knowledge-base of the system and which is relevant to (i.e. is an important element of) the produced output.” Although this framework represents an interesting attempt to combine scientific formalization and cultural production based on Boden’s philosophical work, it is inadequate to capture the different aspects of creativity for reasons similar to those mentioned above.

In addition, certain AI practitioners may pay special attention to the architecture of a system — not in terms of its efficiency, but regarding how well it carries out a particular cognitive process at the functional account. These practitioners typically come from a branch of AI which seeks to gain insights of human cognitive processes from computational modes, and/or vice versa. For instance, the various AI systems coming out of Schank’s group, as part of the renowned Yale School, faithfully implement the notion of “dynamic memory” [185], a cognitive psychology theory explaining how humans use their memories in various tasks.

So far, we have seen that the AI analysis framework concentrates on the operation of computational systems, including the knowledge representation, amount of domain knowledge, algorithmic complexity and performance. Given its focus on the system itself, this framework provides a handy toolset to analyze the algorithmic complexity of intentionality. However, the omission of system authors and users in the framework makes it difficult to address other criteria of intentional systems — process opacity, human-like coherence, and authorial intention. More issues arise when we take a deeper look at the architecture level. As I will argue later in Chapter 3, how well a system architecture models a cognitive process depends simultaneously on the system’s operation as well as the system authors’ narration of it. With these issues in mind, we move on to the existing HCI analysis framework with the relationship between human and computer systems as its primary concern.

2.3.4.2 *An HCI Approach*

A human computer interaction (HCI) specialist may approach a system from a very different angle. He will not pay close attention to how the system *computes* the results, but rather how it *interfaces* with its human users. Evolving within the “computer as a tool” perspective (Section 2.1.2), the field of HCI focuses on the question of how well the tool adapts to the human, particularly its users’ needs and capabilities. Initially growing out of the tradition of human factors and contemporary concerns that computer programs were usually designed without taking the users into serious consideration, HCI promotes human (user) centered design to enhance human productivity through such concepts and design principles as “affordance,” “natural mapping,” “usability,” and “feedback” and avoids execution and evaluation “gulfs” [161]. This goal is often broken down into the subsequent questions of, “How easily can the user *learn* to use the program?” “How *flexible* is the system in supporting the different ways of interaction preferred by different end-users?” and, “Is the system

Table 4: HCI Principles of Usability [40, pp.259-273]

Learnability Principles	Flexibility Principles	Robustness Principles
Predictability Synthesizability Familiarity Generalizability Consistency	Dialog initiative Multi-threading Task migratability Substitutivity Customizability	Observability Recoverability Responsiveness Task conformance

robust enough to provide support for successful achievement and assessment of the users’ tasks?” (Table 4).

Before he looks carefully at either Lewis’ or Cope’s system, the first thing that the HCI specialist would be interested in is its users. Who are the users? What are their characteristics? Are they professional musicians or normal listeners? How much exposure do they have to classical or jazz music? What are their age ranges and income levels, and so on? Unless the HCI specialist knows the answers to these questions, he cannot decide whether the system is usable to *these users*.

Next, he may attempt to identify several benchmark tasks that the systems would need to perform. These tasks are chosen from a set of core tasks that the system supports and are fundamental for the users. As argued below in the next section, it is not easy to define the user or choose such tasks for intentional systems. The term “task” entails an instrument-view of the system, which is incongruous with the original premise of *Voyager*. Some may argue that the author’s intention does not need to be considered in an HCI analysis, but even leaving out authorial intent does not solve our specialist’s problem completely. New issues arise: who is the user of *Voyager*? The audience who is listening to the performance? The human jazz musicians who are playing along with the system? Or, George Lewis himself? If we assume all of the above are legitimate users of the system, then what are the typical tasks of each of these user-groups, and how can a single system be simultaneously helpful and usable to such a diverse range of people with different agendas?

Table 5: Comparison between the AI and HCI Analyses

	AI Analysis	HCI Analysis
Perspectives	Stand-alone Thinking Machine	Computer as a Tool
Subject	System’s internal process	system’s interface
Value	System performance	User’s performance
Measurement	defined by author	defined by studying the users
System Intentionality	A technical property	A result from the interface

Assuming that the HCI specialist somehow resolves all the above difficulties, a more puzzling question awaits him. A core component of HCI evaluation is to measure how “natural” and “easy” a particular system is to use so that the users can increase their productivity. Both interpretive and empirical evaluation methods essentially focus on the question of the performance of the user using the system under certain conditions. Different from the AI definition of performance, here the term applies to the degree to which the users’ tasks are augmented by the specified system. Based on the nature of the benchmark tasks, performance can be measured in terms of speed (time needed to react and/or complete), accuracy (errors, the ratio between hits and misses), production (number of tasks completed), and scores (number of points earned), etc [40]. Although these metrics are useful for certain computer programs that are used to enhance productivity, such as *Photoshop* or tax-filing software, they are not quite applicable to either of the two musical systems of interest to us.

2.3.4.3 *Limitations and Unanswered Questions*

We have just gone through a rough overview of how the fields of AI and HCI, both closely related to intentional systems, offer frameworks for analysis of the two above mentioned computer music systems. AI analysis, continuing the stand-alone-thinking-machine perspective of computing, is deeply invested in a system’s algorithmic complexity and the expressivity of knowledge representation. By contrast, HCI evaluation methods blackbox the internal processes of the system while focusing on its interface with the user. Rooted in the computer-as-a-tool perspective, HCI is more interested

in how the system augments the user’s performance, normally in the form of productivity rather than the performance of the system itself. Compared to the AI analysis, the HCI measure of performance is defined based on the study of the users and the context, instead of being defined by the author(s) as is the case with the AI system (Table 5).

These two analytic schemas offer in-depth knowledge of either the formal properties of a system or the ease with which human users can adopt them; however, each exhibits considerable limitations when applied to intentional systems. As we have seen, AI analysis typically completely ignores context and focuses on computational systems only in isolation without considering either the author or the user of the system. Although process complexity is a necessary condition for intentional systems (Section 1.2), it is not sufficient by itself to form intentionality. Following the practice of natural sciences, the “content” of the system matters to AI only in terms of its formal representation. In other words, whether Cope chooses to model Mozart or a first-year composition student is not relevant. HCI, on the other hand, provides certain vocabulary and methodology to consider users and the context of interaction; however, the field’s current productivity-oriented value system prevents it from being further helpful to the study of intentional systems. As we have seen from the previous section, without a clear definition of the users’ needs that should be fulfilled by the particular system, the HCI analysis cannot be properly applied.

Neither of the two evaluation schemas tells us much about where or how system intentionality is formed. Neither schema explains, for instance, how authorial intention effects the users’ perception of system intentionality — how Lewis’ political stance impacts how his audience sees *Voyager*; how listeners’ perceptions of the intentionality of *EMI* are structured by such provocative names of the albums as *Virtual Mozart*. Nor do these schemas take into consideration the larger discourse about these systems that happens *outside* of the machine operation and outside of

the human’s direct interaction with them. Covering *EMI*’s famous “Mozart’s 42nd Symphony,” a reporter wrote:¹³

MUSIC LOVERS who attended a concert at the University of California at Santa Cruz in April heard the college orchestra put their hearts into Mozart’s 42nd symphony. The work features all the glorious harmonies and elegant flourishes one expects from Mozart, though it perhaps lacks some of the genius of other late works such as the 41st symphony, the famous Jupiter. But there is a reason for this lack of *je ne sais quoi*. In the 207-year interval between writing the 41st symphony and the 42nd, Mozart was busy doing more decomposing than composing. [99]

Many curious listeners of *EMI* first encounter such media coverage about *EMI*, before they listen to the album. Does it possibly impact their perception of the system? If so, in which ways does the impact occur? These questions will be answered in the next chapter.

2.4 *AI-based Interactive Narrative*

The last component of the theoretical framework pertains to a specific application of the theory of intentional systems, namely how to design intentional systems in the domain of interactive narrative. Chapter 6 presents our interactive narrative project *Memory, Reverie Machine* (MRM), as a case study of the intentional system design principles proposed in Chapter 5. This section provides the necessary background of interactive and generative narrative needed to understand the design and contribution of MRM. I first introduce the expressive AI practice as the overarching framework for my own design and construction of intentional systems. Harrell’s approach to computational (interactive and generative) narrative and the GRIOT system is briefly discussed as both the theoretical and technical foundation for the *MRM* project.

¹³Mozart composed 41 symphonies in total during his lifetime.

2.4.1 Expressive AI Practice

Expressive AI is a new interdisciplinary approach to AI proposed by the computer scientist Michael Mateas [135, 136]. It combines AI research agenda with artistic practice by its focus on interpretation and authorship. Different from both GOF AI and interactionist AI, whose ultimate goal is machine intelligence despite their distinctive approaches described above, expressive AI foregrounds its *interpretive affordance*, which supports the reading of both the system’s underlying operation and user interactions. Expressive AI’s criteria of success diverge from traditional AI research on performance and efficiency and are based on whether the AI practitioner/artist can establish a specific poetics that engage the audience in specific processes of interpretation.

Unlike many other (interactive) art practices, expressive AI practice also pays attention to what is *inside* the AI-based system. Different internal architectures of an AI system afford different possibilities that a practitioner can use to inscribe her authorial intention on the machine — a concept Mateas defines as an *authorial affordance* of specific AI architecture. The system architecture itself, as exemplified in Mateas’ AI-based interactive art piece *Terminal Time* [136], can contribute to the meaning of the piece. For instance, one may take a less generative approach to implement *Terminal Time* and still exhibit similar results. The conceptual strength of the piece, however, will be substantially weakened, according to Mateas, because the interpretive and architectural concerns are no longer aligned. As Mateas claims strongly, “[a]n artist engaging in expressive AI practice will be able to build works with a depth, a richness, a sophistication that can’t be achieved without this simultaneous focus on meaning making and machine structure.”

Expressive AI serves as the underpinning of both my theoretical inquiry and expressive practice in multiple ways. Most importantly, it demonstrates that the attendance to a system’s technical specifications, including system architecture, data

representation, and algorithmic choices, is not necessarily orthogonal to humanistic concerns of meaning and expression. In fact, as the rest of the dissertation will illustrate, these technical insights are relatively unexplored an area for humanistic and critical study of the relationship between human and computational technologies, such as intentional systems. Second, my own interactive narrative practice, presented in Chapter 6 as a case study, aligns closely with expressive AI in its simultaneous focus on authorial intention and user interpretation through AI-based techniques.

2.4.2 Computational Narrative

Computational narrative is a major application area for the insights regarding intentional systems provide by this dissertation. In particular, Fox Harrell’s approach to computational narrative has been influential [81, 83]. His major concerns are exemplified by his *GRIOT* system, which is the foundation of *Memory, Reverie Machine* project both in terms of technical implementation and our approach to computational narrative [80]. This subsection offers a brief review of this perspective, as laid out in Harrell’s dissertation [81], while emphasizing its influences on *MRM*.

Harrell’s framework on computational narrative engages works with the following characteristics:

generative content, semantics-based interaction, reconfigurable narrative structure, and strong cognitive and socio-cultural grounding. A system that can dynamically compose media elements (such as procedural computer graphics, digital video, or text) to result in new media elements can be said to generate content.

He further elaborates the emphasis on meaning and semantics in his formalization and computational methods:

Semantics-based interaction here means that (1) media elements are structured according to the formalized meaning of their content, and (2) user

interaction can affect content of a computational narrative in a way that produces new output that is “meaningfully” constrained by the system’s author. More specifically, “meaning” in *GRIOT* indicates that the author has provided formal descriptions of domains and concepts to either annotate and select or generate media elements and subjective authorial intent.

Meaning can also be reconfigured at the level of narrative discourse. The formal structure of a computational narrative can be dynamically restructured, either according to user interaction, or upon execution of the system as in the case of narrative generation. Discourse structuring is accomplished using an automaton that allows an author to create grammars for narratives with repeating and nested discourse elements, and that accept and process user input. Appropriate discourse structuring helps to maintain causal coherence between generated blends. Strong cognitive and socio-cultural grounding here implies that meaning is considered to be contextual, dynamic, and embodied. The formalizations used derive from, and respect, cognitive linguistics theories with such notions of meaning. Using a semantically based approach, a cultural producer can implement a range of culturally specific or experimental narrative structures.

Technically, *GRIOT* implements the core elements of conceptual blending theory using Josephe Goguen’s computer science theory of algebraic semiotics [71]. Details can be found in their joint work [72].

Harrell states that his goal with *GRIOT* was quite different from passing a type of Turing test for creative competence, as described earlier in this section. Instead, he designed it to provide a technical framework for humans to provide rich content, and narrative systems created with *GRIOT* are meant as cultural products themselves (as opposed to instances of output of such poetic systems). It utilizes cognitive

science models, informed by the skepticism of logical AI modeling of human thought in the cognitive linguistics enterprise [119], toward expressive ends that are often literary. The *Memory, Reverie Machine* quickens this interest in terms of subject matter and interaction model as well. As shown in Chapter 6, it is used to explore the meaning and limitations of machinistic thought through fiction, while critiquing its outmoded notions theoretically and technically. Finally, the notion of generating memories, reveries, and cascading sequences of remembered events is reinvigorated by work such as conceptual blending theory and novel techniques informed by AI-based interactive narrative.

In summary, this chapter presents my synthesized and integrated theoretical framework drawn from a broad range of areas. It is organized around four core questions, including 1) the role of intentional system in relation to human beings, 2) the possibility of system intentionality, 3) ways to “read” intentional systems, and 4) approaches for designing intentional systems in the domain of interactive narrative.

The first section is organized around the major perspectives on the role of computers in our society in the history of computing. The stand-alone AI view, instrumental view, medium view, and self-reflexive view emerged in different historical moments of computing, all of which have their respective influence on intentional system. Next, I traced the lineage of the notion of system intentionality, first through the ontological debates of AI in the area of philosophy of mind, and then through various other theoretical developments that portray a broader notion of system intentionality. The following Section 2.3 offers a set of theories, including the field of hermeneutics, software studies, cognitive semantics, AI and HCI frameworks of system analysis, that provide useful theoretical and methodological tools for close reading of intentional systems. Finally, the last section gives a brief account of the existing practices that influenced my own work in designing and constructing of an intentional system in the

domain of interactive narrative. Prepared with this synthesized framework, we are now ready to tackle the first core research question — How does intentionality form in intentional systems?

CHAPTER III

THE AI HERMENEUTIC NETWORK

What makes any hardware or software component an agent is precisely the fact that one has chosen to analyze and control it in these mental terms.

— Yoav Shoham, “Agent-oriented programming” [196], 1991

Dennett’s intentional stance implies that system intentionality is derived by the users who want to predict the behavior of the system they are interacting with. This section will extend Dennett’s argument and add system author’s narration of intentionality as an equally important component that gives rise to system intentionality. Philip Agre [2], one of the most well regarded thinkers in dialogue with both the technical and philosophical communities, explicitly points out the discursive nature of AI practice. Based on his first-hand experience as an AI practitioner, Agre argues that “the purpose of AI is to build computer systems whose operation can be narrated using intentional vocabulary” [2], such as “reasoning,” and “planning.” Most AI practitioners, according to Agre, find the ontological debate of whether these systems really exhibit these phenomena “an annoying irrelevance.” To them, what really matters is not the “vague” issue of what the words really mean, but “the seemingly precise issue of how they can be defined in formal terms that permit suitably narratable systems to be designed” [3].

The action of narration, however, is not as straight-forward as it may seem. In AI, the practitioner’s intentional narration of her system first needs to be differentiated from a fairy tale that an adult constructs to cheer up an unhappy child. In the

latter scenario, the adult intends to achieve a specific goal by constructing a narrative that she knows is counter-factual. The intentional narration of AI, on the other hand, is so seamlessly integrated into the practice of AI that most practitioners participate in it without being consciously aware of it. As argued later, the dichotomy of “factual” and “counter-factual” is derived from a romanticized notion of science. Second, the intentional narration is manifested not only through the literature (e.g., interviews, technical publication, personal communications), but also through the technical construction of AI systems, including the choices of function, algorithm, system architecture, as well the representation. Third, narrated system intentionality needs to be communicated to the users of intentional systems, who subsequently interpret the system and develop their own intentional narrations of it. Most importantly, the narrations and interpretations are interconnected through a complex network of meaning exchange and negotiation between the authors and the users of intentional systems in the larger social and cultural context.

The above paragraphs explain the main idea of my overall framework of an AI hermeneutic network, which emphasizes that system intentionality arises from a hermeneutic process, which incorporates three types of elements: the system author’s discursive strategies, the system’s discursive machine, and the user’s hermeneutic reading. Part of my major contribution is to critically analyze the discursive nature of system intentionality from the vantage point of an insider-outsider of AI. Similar to an ethnographer who, when living in a different culture, gains “inside” information about a group of people she is studying, my approach to AI is based on my own experience as an AI practitioner. At the same time, as the ethnographer is essentially an “outsider” and inevitably makes sense of this culture through her own lens, my critique and approach to AI is informed by the theories of the humanities and social sciences.

Figure 13 illustrates the distinction between this new framework, the AI analysis

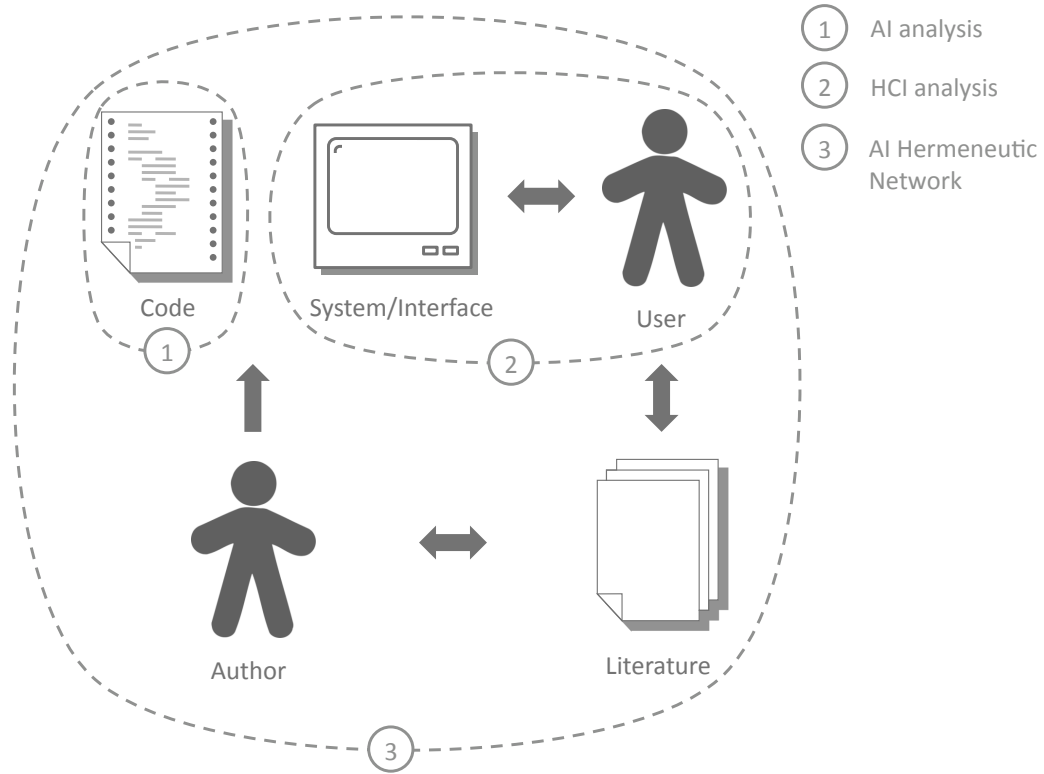


Figure 13: The AI Hermeneutic Network

and the HCI analysis schema. As described earlier in this chapter, the scope of the traditional AI analysis is within the code. The algorithm and knowledge representation are studied extensively in terms of their effectiveness and expressivity. The HCI analysis is primarily concerned with the interface of a system. The central concern is how user-friendly the system is and how well it matches the needs of its users. In contrast, the AI hermeneutic network takes into account all the above elements as well as the discourse of the system in the form of literature.

This chapter presents the theoretical framework of the AI hermeneutic network. First, Section 3.1 proposes a textual approach to analyze intentional systems, which addresses some of the issues of the existing AI and HCI analysis, discussed previously in Section 2.3.4. Next, Section 3.2 draws from the lessons from existing analysis of Alife, and identifies the discursive strategies that system authors deploy at the level of source code, presentation/interface, and literature about the system. Next, Section

3.2.6 briefly discusses the user’s hermeneutic reading of the system. Finally, Section 3.3 expands on Figure 13 by inserting new connections between different elements and extends it to a broader social context.

3.1 Intentional Systems as Text

The limitations and difficulties we have encountered when applying the existing AI and HCI frameworks to intentional systems suggests that system intentionality is not simply the result of well engineered algorithms and knowledge representations, or the “natural” interaction between user and the system. The four criteria of intentional systems call for a different model that better addresses the correlation between user, author, and the system. In this section, I propose a textual-approach, that is, to study intentional systems as an artifact that participates in the meaning-making process between its creators and users. This section first explains the benefit of analyzing intentional systems as texts and identifies certain traditional textual analysis approaches that are useful for locating the formation of system intentionality. Then, it makes explicit the reason why one particular historical approach, hermeneutics, can be useful for our purpose.

3.1.1 What is *Text*?

The meaning of *text* has been broadened tremendously by such contemporary theoretical developments as structuralism, post-structuralism, literary theory, semiotics, cultural studies, and modern hermeneutics. Especially under the influence of Roland Barthes’ work, among others, the term refers not only to written and print-based forms of discourse but also to a much broader range of cultural productions [9, 7]. Novels, paintings, photographs, fashion, TV programs and advertisements can all be analyzed as “Text,” despite their different modalities and representations, because they rely on the same semiotic structure in the meaning-making process, or *codes* [8], as written discourses.

Many approaches and theories have been developed about what constitutes as texts and how to study them. For instance, Barthes' work develops Saussure's dream of semiology, a general science of all languages and social signs [184]. Building on Saussure's linguistic theory, theorists such as Claude Lévi-Strauss and Barthes developed a structuralist approach in order to understand the process of meaning-making through universally repeated patterns. Barthes's well-known semiological analysis of cultural "myths" [7] reveals the underlying structural pattern that allows certain dominant meaning-making to appear "natural" and "self-evident." For Barthes, everything is meaningful and symbolic and therefore can be treated as text. Michel Foucault[59, 60], on the other hand, believes that "discourse" should be the central unit of analysis because everything is mediated through power relationships and language. More recent analyses tend to move away from the semiological/structuralist approach. Post-structuralist theorists such as Derrida, Deleuze and Guattari adopt a more open and fluid approach on the basis of the incompleteness and impreciseness of language and meaning-making.

The primary motivation for analyzing intentional systems as texts here is not to propose coherent sign systems, or to engage in a discussion of the social struggle around their production and consumption, even though it is an important topic. Instead, a textual studies approach for intentional systems highlights the discursive and interpretive nature of these technical systems, which is still relatively unexplored under the current AI and HCI paradigms described above. Seeing intentional systems as texts thus opens up new possibilities to consider them as part of the meaning negotiation process between the author and the user, a central topic that has long been part of textual studies. Although traditional approaches to text (Section 2.3.1) assume that the meanings encoded within texts by the authors can be objectively retrieved by readers without any distortion, given appropriate methods of interpretation [92], many contemporary theorists believe that meaning is made by the reader/viewer

[200]. The AI hermeneutic network model introduced in the next section subscribes to a more balanced view. As German philosopher Manfred Frank [62] proposes in an essay on the literary text, texts support a complex communication process in which both author and reader actively create, shape, and reconstruct meanings. This echoes the even broader notion of dialogic meaning posited by the Russian philosopher and critic Mikhail Bakhtin, who suggested that language is understood as dynamic, contextual, intertextual, and relational [100]. Therefore, acknowledging the textuality of intentional systems allows us to engage a range of socially situated methods that have been developed outside the fields of AI and HCI. For instance, the well-studied social and cultural impact of the meaning-making process can be used to explain different readings of intentionality of the same system (e.g., *ELIZA*) for different users, a phenomenon described in Section 1.2.

3.1.2 A Hermeneutic Framework

Out of the wide range of textual analysis theories and methods, hermeneutical analysis (a brief overview of hermeneutics is available in Section 2.3.1) provides a unique vantage point for studying the formation of system intentionality and integrating other more contemporary theories for our purpose. Compared with its modern counterpart theories, hermeneutics (particularly classical hermeneutics) emphasizes the role of authorial intention. In comparison, most contemporary theories see authorial intention as irrelevant to the meaning of the text. This position is famously captured by Barthes' announcement of "the death of the author"[9]. Saussurean linguistics [184] and its closely related structuralist approach to text also keeps the author out of any core analysis. However, we have to take into account the context from which these theories emerged. The assumption that the author-intended meaning is the only valid interpretation has dominated classical hermeneutics and textual interpretation for centuries. Barthes' denunciation of authorial intention should not be understood

literally, but instead as his strategy to call attention to the plurality of readers' interpretation, which is long neglected and discarded in this tradition. The development of AI and intentional systems did not share the same historical path as text.

In the modern practice of computing, the author's intention is often suppressed in the name of science and the users' needs. AI practitioners often see themselves as objectively and neutrally following a scientific procedure, whereas designers who subscribe to human centered design principles prioritize their designs according to what the user wants, rather than their own design taste. However, the meaning exchange model between the author and user reveals the implausibility of the valorized notion of a neutral and objectified author. As Winograd and Flores observe, the system programmers do not have absolute freedom to design a system because "[t]he programmer acts within a context of language, culture, and previous understanding, both personal and shared"[217, pp.97]. Downplaying the role of human system authors, therefore, risks omitting these contexts and the strategies deployed by the authors in the production of intentional systems. In other words, restoring the role of the author, as opposed to downplaying it, is the first step to studying intentional systems as texts.

Acknowledging authorial intention and their subsequent discursive strategies also helps to prevent the techno-centric view in technical disciplines that system intentionality arises from the computer itself. In fact, as the next section argues, a hermeneutic-based approach helps us understand how system intentionality is *narrated* by the system's authors and interpreted by its users. In this regard, the spirit of hermeneutics helps us to compensate for the limitation of the current AI and HCI framework (discussed in Section 2.3.4) and provides a promising new perspective for understanding the formation of system intentionality.

It is worth mentioning that, despite its primary focus on author intended meanings, hermeneutics still has certain room for different reader interpretations. A good

interpreter not only retrieves the original meaning embedded in the text by its author, but also applies it to her current social and cultural context so that the original meaning becomes relevant. In hermeneutic study of biblical text, the ability to apply the divine message in historical texts to a different audience in a different society is a crucial skill for an interpreter [208]. Of course, the flexibility and plurality in the practice of hermeneutics is limited compared to the flexibility in later theories, such as the post-structuralist approaches. This limitation can be addressed by integrating such theories with the classical hermeneutics framework. However, as the rest of this chapter and the following one will show, the primary emphasis of this dissertation is a system’s authors’ narration of system intentionality.

3.2 Constructing the Discursive Machine

Faced with the difficulties of pinning down a precise definition of agenthood, a key concept in the AI domain of multi-agent systems (MAS), Yoav Shoham wrote:¹

An agent is an entity whose state is viewed as consisting of mental components such as beliefs, capacities, choices, and commitments. These components are defined in a precise fashion, and stand in rough correspondence to their common sense counterparts. In this view, therefore, agenthood is in the mind of the programmer: what makes any hardware or software component an agent is precisely the fact that one has chosen to analyze and control it in these mental terms. [196, pp.329]

Shoham’s comments correctly locate agenthood as the result of the author’s willingness to adopt intentional vocabulary as her lens to analyze and build systems, rather than as a property of these systems. However, he overlooks the impact of

¹Agenthood is a highly intentional term. It is commonly used in the research area of software agents and multi-agent systems, but its precise definition has not yet been agreed upon. It generally refers to that the agent has its own personality and/or history of interaction [46].

these terms on the ways they structure the AI researcher’s practice. This section argues that what Shoam calls the “rough correspondence” between the formally defined AI intentional terms to their “common sense counterparts” constitutes an essential component of AI and intentional systems — the discursive machine. A discursive machine contains the discursive strategies that a system author deploys to narrate system intentionality. This machine is so embedded as part of the AI practice that most practitioners are not consciously aware of its existence and influence. Among the few AI practitioners who notice its operation, some see it as the epidemic of true scientific research and try to eliminate it at all costs. This section concentrates on the usage of intentional vocabulary. Initially built on Agre’s observation on key AI intentional terms, it then broadens the scope of intentional vocabulary to include a wider range of discursive strategies, including definitions, algorithms, system architecture, and literature, as well as visual vocabulary (e.g., images of anthropomorphized agents). Again, the pervasiveness of the discursive machine should not be seen as an obstacle towards the “real” AI. Instead, as argued earlier, it is a crucial part of the meaning-making process between the author and user, in which system intentionality is formed.

This section first draws from Katherine Hayles’ existing analyses of Alife, in which she argues that the various level of narratives constructed by Alife researchers are constitutive to the research area. Extending Hayles’ study to incorporate the actual technical practice of AI, Section 3.2.2 highlights the significance of intentional vocabulary to the field by re-examining AI practitioners’ related observations through the lens of actor-network theory. Next, Section 3.2.3 to Section 3.2.5 illustrate system authors’ various deployment of discursive strategies to narrate system intentionality with examples at the source code, presentation, and technical literature level of their systems.

3.2.1 Lessons from Alife: Discursive Strategies

Before introducing the AI hermeneutic network framework, it is useful to position my approach with N. Katherine Hayles' [87] work on Artificial Life (also referred to as "Alife," "alife," and "AL") for two reasons. First, Alife computer programs are very similar to intentional systems in that they (are intended to) display phenomena commonly associated with real life forms, such as cells and birds, not machines. Second, Hayles's study of the Alife programs and research area focuses explicitly on the discursive strategies that Alife practitioners deploy. After a brief account of this existing work, this section discusses how my approach builds upon Hayles' work and yet differentiates itself by underlying the technical practice as discursive as well.

As its name suggests, Alife is an area that is closely related to AI with researchers working in both areas. The founder of the interactionist AI paradigm, Rodney Brooks, for instance, is also closely associated with Alife. As stated in Section 2.3.1.3, the interactionist AI paradigm turns its back on the classical AI (GOFAI) approach of modeling intelligence through formal rules and focuses instead on the embodied interaction between the system, called an agent, and its environment. These embodied agents, often compared to primitive life forms such as ants and cockroaches [20, 21], are seen as one kind of artificial life form in Alife. If AI is the study of computational systems in relation to intelligence, Alife's focus is then on systems as life forms.

Startled by the Alife researcher and evolutionary biologist Thomas S. Ray's assertion that "[t]he object of an AL instantiation is to introduce the natural form and process of life into an artificial medium" [175, 180], Hayles realizes that Ray's rhetoric exemplifies Alife's underlying message that computer codes or hardware may be living organisms. Hayles goes on:

How is it possible in the late twentieth century to believe, or at least claim to believe, that computer codes are alive — and not only alive, but natural? The question is difficult to answer directly, for it involves

assumptions and ideas circulating throughout the culture. In view of this complexity, the subject is perhaps best approached through indirection, by looking not only at the scientific content of the programs but also at the stories told about and through them. These stories, I will argue, constitute a multilayered system of metaphors and material relays through which “life,” “nature,” and the “human” are being redefined. (pp.224)

In a subsequent study, Hayles examines the “narratives” at three levels. The first level of narrative includes “representations, authorial intention, anthropomorphic interpretation” of a specific Alife computer program written by Ray and others called *Tierra*. Compared to Richard Dawkins’ work *The Selfish Gene*[34], Hayles emphasizes that Ray does not simply “attribute” human agency and intention to genes as Dawkins. Instead, Hayles is “working with artificial systems designed by humans *precisely so the ‘creatures’ would be able to manifest these qualities.*” A crucial component of the manifestation is constructing the narratives, so that they are tightly interwoven into the operations of the program. As a result, Hayles observes, “the program operates as much within the imagination as it does within the computer.” For instance, she points out that using such terms as “mother cell,” “daughter cell,” “ancestor,” “parasite,” and “hyperparasite” does more than set up parallels between the computer program and the living systems; these terms also “reveal Ray’s intention in creating an appropriate environment in which the dynamic emergence of evolutionary processes could take off” (pp.227).

At the second level, narratives are effective tools that Alife practitioners mobilize to seek legitimization of their research area within theoretical biology. Different from the ones told at the first level about individual programs and hardware, the narratives here are about the research area of Alife itself. In order for their work to be accepted in theoretical biology, Alife researchers need to position these programs as *life-as-it-could-be*, compared to which the traditional biology of *life-as-we-know-it* is a

special case. Some researchers, Hayles notices, compare Alife simulations to synthetic chemicals, which freed the early study of chemistry from the confinement of naturally occurring elements and components. The goal of this analogy then is to argue that Alife is equally important to theoretical biology, for it is not a model or simulation of carbon-based life forms, but a second instance of life.

At the third level are narratives about the present and future of terrestrial evolution forms. These are constituted by speculating on the relation of human beings to the “creatures” created by computer software or hardware. According to Hayles, rather than being merely the measure to judge the success of AI systems, human intelligence in Alife is itself reconfigured in the image of the evolutionary processes modeled in Alife. In other words, the machine in Alife does not simulate the human, but instead becomes the model for understanding the latter (pp.224-239).

Hayles’ analysis of what she calls “the nature and artifice of Artificial Life” is revealing and relevant to intentional systems, for system intentionality, as I will argue below, is deeply discursive as well. The methods she uses directly inspire my construct of the *AI hermeneutic network* because they reveal how the “aliveness” of the Alife programs is narrated by Alife practitioners through the “stories” they construct. However, I do not intend to repeat the above work here, but rather to connect it more closely to the actual practice of AI and its related fields by addressing the following aspects that have not been taken into serious consideration.

First, Hayles stresses the importance of the rhetorical strategies that Alife practitioners adopt. However, these rhetorical strategies are taken mainly from the literature (e.g., papers, talks, interviews, and interpersonal communications) not from technical *practice*. As I will show in the next section, these technical practices of intentional systems are not less discursive than the literature. In fact, they are part of the AI technical practice without which no AI system can function.

Second, Hayles’s analysis implies that the leading Alife practitioners knowingly

constructed these narrative of the systems and the field, *in order to* be accepted by the larger intellectual community of theoretic biology. The next section will illustrate the ways in which these discursive strategies are so entrenched in the practice of AI that they are rarely acknowledged or questioned by most practitioners in the field.

In other words, my approach extends Agre and Winograd’s main philosophical critique of AI with a textual analysis approach informed by Hayles’ work. Both Agre and Winograd are concerned with the mentalist fallacy embedded in the symbolic AI paradigm, Winograd & Flores [217], and Agre [2] subsequently points out that the actual practice of AI is inadequate to capture the basis of human existence and understanding. Although my AI hermeneutic network framework echoes a lot of Agre and Winograd’s observations of the field of AI, one of the differences is that I do not see the discursive and interpretive nature of the technical practice as an obstacle to achieving intelligence and intentionality.

3.2.2 The “Epidemic” of Intentional Vocabulary

Throughout the history of AI, new technological innovations have brought in such intentional vocabulary to the field as “reasoning,” “planning,” “learning,” “strategizing,” and “creating.” These intentional terms are so pervasive in AI that it is impossible to talk about any algorithms or systems without using them. To the practitioners in the field, these terms have very specific meanings relating to specific methods, which are only “roughly correspondent” to their commonsense meanings. This section calls into attention the close interplay between the formal and commonsense meanings of these intentional terms, based on the observations of AI practitioners including Agre and McDermott. Different from their perspectives that intentional vocabulary is an epidemic whose usage should be minimized, this section argues that the intentional terms are an constitutive component of AI practice.

The connection between the precise formal meanings and their vague common

sense counterparts of the intentional vocabulary is what Agre calls the “dual character of AI terminology”:

When a running computer program is described as planning to go shopping, for example, the practitioner’s sense of technical accomplishment depends in part upon the vernacular meaning of the word.... On the other hand, it is only possible to describe a program as ‘planning’ when ‘planning’ is given a formal definition in terms of mathematical entities or computational structures and processes. [3]

In this sense, the elasticity of these key terms to be simultaneously formal and vague is strategic. In the meantime, Agre denies that using intentional vocabulary is a conscious deception. Although it broadens the application of AI, this practice is also “self-defeating” because these elastic terms inevitably link AI to a much larger discourse based on reflections of their vague meanings. The consequence is that AI practitioners “find it remarkably difficult to conceptualize alternatives to their existing repertoire of technical schemata” [3].

Drew McDermott [140] made a very similar, but more radical, observation on this issue in the mid 1970s. Criticizing the relationship between the formal and vernacular meanings of intentional vocabulary as “wishful mnemonics,” McDermott saw it as “a major source of simple-mindedness in AI programs.” Identifying the use of these intentional mnemonics in a wide variety of AI systems ranging from the *General Problem Solver* (*GPS*) to language “translation” systems, McDermott warns AI practitioners that the epic of “contagious wishfulness” is misleading, most prominently to the practitioners themselves. Instead of naming their programs “UNDERSTAND” or “THINK,” all disciplined programmers, he urges, are to avoid the trap of wishful mnemonics. What they should do is to refer to their program as “G0034” and see if they can still convince themselves or anyone else that G0034 implements some part of understanding.

McDermott succinctly traces the usage of mnemonics to the earlier structured programming paradigm. When a system is implemented under this paradigm, it is programmed “entirely in terms of the purposes at one level before implementing them by the most convenient of the (presumably many) alternative lower-level constructs.” For example, a Google-like program may be defined by several high-level functions (e.g., “query,” “sort,” and “display”), each of which is defined by its specific purpose. At the next step, the high-level functions are subsequently broken down into the more concrete functions, and the most proper algorithms are then used to implement the pre-defined purposes. This is the step where the quick-sort algorithm is chosen to instantiate the sort function. This top-down programming approach relies on a close match between the purpose of a function through its name and its operation. In other words, the meaning of “sort” is precisely defined and widely shared by the programmers. The situation is quite different in AI. McDermott argues that:

If a researcher tries to write an ‘understanding’ program, it isn’t because he has thought of a better way of implementing this well-understood task, but because he thinks he can come closer to writing the *first* implementation. If he calls the main loop of his program ‘UNDERSTAND’, he is (until proven innocent) merely begging the question.

This brief observation provides a way for us to understand Agre and McDermott’s denunciation of mixing the vernacular and formal meanings in AI practice. A deeper reason at work here is the romanticized view of science. According to this view, science is an objective reflection of nature and exists independent of its practitioners and society. The structured programming paradigm represents a top-down engineering approach for specifying a program using the “purpose” of each function. With this level of detail, programmers can implement and fulfill these purposes using methods commonly understood and agreed upon within the community. This process is

similar to instantiation of a conceptual architectural blueprint to detailed construction diagram with specifications of the construction material needed, and eventually to a real building. Each level of detail defines the modules' purpose, which is realized with more concrete details at the next step.

Writing an “understanding” program is different from instantiating the functions inside a data flow diagram (e.g., a sorting function) not merely because of the lack of understanding of how to do so, as McDermott argues. A deeper issue here is the discursive and declarative power that the AI practitioner possesses to name her program “UNDERSTANDING,” something that the programmer writing top-down structured programming code does not actually have. The active role of scientists (and engineers) in inscribing and narrating the results of their experiments has long been neglected for the sake of “objectivity” in scientific practice, especially experimental natural science. Under the rhetoric of this romanticized notion of science, its disciplined practitioners are neural devices immune to their own “subjectivity.” However, the dilemma here is that nature cannot speak directly. As Latour [122, pp.70-74] asserts, part of the scientist’s mission is to be the *spokesperson* for what is inscribed by her instruments. In a hypothetical scenario, Latour depicts a reader who is determined to see the proof of an argument about endorphins from a scientific paper. She visits the author of the paper and is asked to read the results of the experiments herself. By illustrating the difficulties the reader goes through to connect the readings from scientific instruments and the argument made in the paper, Latour effectively demonstrates the power of the spokesperson, not because of her own wishes, but because of her task to transmit what nature really is. To summarize Latour’s point, no scientific practitioner/author can be disassociated from her claim.

What Agre and McDermott did not acknowledge is the spokesperson role of AI practitioners. Certainly, the romanticized notion of science makes it more comfortable to assume that the operation of a system is independent of its authors. For this

reason, McDermott suggests all functions should be named in the style of “G0034.” However, even McDermott himself notices the formidable difficulty of his proposal. An assembly-like AI program without the narration of its author is like a reading in an instrument waiting for the scientist to be its “mouthpiece.”

The examples of “UNDERSTANDING” and “G0034” are both extreme. In most cases, AI practitioners are simultaneously the executor and narrator for their systems. On the one hand, the formal meanings of many key intentional terms have been established and followed by AI practitioners in their systems. Simply naming a program “PLANNER” does not automatically legitimize it as a planner in the AI sense. In this sense, an AI practitioner (only) executes the conventions and methods associated with these terms. On the other hand, the operations and narrations of the systems need to be narrated. Similar to the Alife researchers in Hayle’s analysis (Section 3.2.1), AI practitioners are different from theoretical scientists who identify certain properties in existing subjects. AI practitioners are, instead, engaged in the task of creating artifacts that manifest these properties, artifacts that did not exist before. The elasticity of the intentional vocabulary hence provides AI practitioners with an effective discursive vehicle that closes the gap between the operation of a system and the properties it is designed to exhibit. I argue that without the glue of intentional vocabulary used in practitioners’ narrations, the empire of AI may collapse. Therefore, the “wishfulness” is “contagious” not because it is “deceptive” but because it is *necessary*.

Intentional vocabulary is one of the most important ways through which system intentionality is narrated. The simultaneous role of AI practitioners as executors and narrators is reflected in their technical practice, detailed in Chapter 4. However, the understanding of “vocabulary” does not need to be limited to just AI key terms such as “planning.” In the following two sections, I will extend Agre’s work and argue that the blend (Section 2.3.3.2) between the formal and vernacular meanings of these

terms also structures the discursive machine through various other forms, at both the code and the interface level.

3.2.3 The Code Level

Agre’s insights on the duality of intentional vocabulary help to settle the debates between the software formalists and software culturalists. As Section 2.3.2 describes, the former are primarily interested in the execution of the code, whereas the latter focus on the cultural significances of it. Agre’s argument suggests that the formal and cultural elements of code cannot be separate completely from each other, because they are ultimately linked together through the elastic meanings of the vocabulary they share.

Michael Mateas [136] reached a similar conclusion in his dissertation. He deconstructs the codes invoked in AI practice into a “code machine” and a co-existing “rhetoric machine,” both of which are necessary for any AI system to function. The former includes the physical processes, computational processes, and complex causal flow specified in the code, whereas the latter is composed of discursive strategies, humanly interpreted computation, and definitions of system progress. Although the code machine handles the execution of the system, according to Mateas, it is the rhetoric machine that allows both the author and the reader of the code to make sense of and narrate what the code does.

This section demonstrates the discursive strategies that AI practitioners deploy to narrate system intentionality in their source code. It involves the choices of both the definitions and the system architecture. Through various examples, I intend to illustrate how source code can embody effective meaning-making strategies and hence I stress the importance of code in the analysis of computational system, both formally and culturally.

```

(defun eliza ()
  (loop
    (print 'eliza>)
    (write (flatten (use-eliza-rules (read))) :pretty t)))

(defun use-eliza-rules (input)
  (some #'(lambda (rule)
    *eliza-rules*)
    (let ((result (pat-match (rule-pattern rule) input)))
      (if (not (eq result fail))
        (sublis (switch-viewpoint result)
          (random-elt (rule-responses rule)))))))

(defun switch-viewpoint (words)
  (sublis '((I . you) (you . I) (me . you) (am . are)) words))

```

Figure 14: Sample code of *ELIZA* from Norvig’s Textbook *Paradigms of Artificial Intelligence Programming* [162] (Emphasis Added)

Function Level Figure 14 shows an excerpt of the source code of a program that re-implements *ELIZA*. The code is from Norvig’s Lisp programming textbook[162], which is widely used among computer science students.² The main functions of the program call another function, whose name is “switch viewpoint.” In a vernacular understanding of the term, to switch someone’s viewpoint means to put that person in the shoes of another person, to adopt that other person’s specific social, cultural, political viewpoint in a specific situation. However, the code of the function “switch-viewpoint” at the bottom of the figure shows that the only manipulation that this function does is to swap the words “I” for “you” and “am” for “are,” etc.

Under a traditional computer science and AI analysis (Section 2.3.4.1), the name of a function is a low-level implementation detail irrelevant to the “essence” of the program, for it does not effect the machine execution of the code. However, in the actual practice of AI, such terms and definitions provide the system authors with *discursive anchors* so that they can narrate the operation of the systems as intentional.

²Lisp is a family of computer programming languages that is widely used for developing AI system. It was initially invented by John McCarthy in 1958.

Describing the operation of the function as “*ELIZA* is switching her view point” generates foreseeably different reactions from the less intentional narration of “*ELIZA* is replacing the ‘You’s with ‘I’s in the sentence.”

Algorithm Level The discursive machine also lends itself to the design and narration of algorithms. In many cases, it is not just the terminologies that are used to define a particular function; the entire algorithm is mobilized to support a coherent intentional narration, as shown in the following example. In the subfield of machine learning, reinforcement learning is a family of algorithms that learn to choose optimal actions to achieve their goals [148]. A reinforcement learning equipped autonomous agent can experiment in its environment in order to learn a control strategy, called *policy*, to choose actions that achieve its goals based on the “rewards” and “punishments” it receives.

One example is a camera sensor-equipped robot whose goal is to dock into a power plug whenever its battery level is low. The robot has a set of actions, such as “move forward” and “turn,” that allow it to achieve its goal. Its goal is defined by a reward function that determines the number of rewards and punishments for each state, locations in this case, that the robot is in. For instance, any state-action transitions that immediately provide a connection to the power plug receive a positive reward, whereas the transitions that cause the robot to leave the room where the plug is get a negative reward or punishment. A reinforcement learning algorithm learns the optimal policy based on the rewards for each action.³

The idea of modifying an agent’s behavior with positive and negative consequences has been explored by the psychology theories of classic and operant conditioning. In the psychological studies of learning and action selection, human and animal subjects are shown to alter their behaviors to obtain positive rewards and avoid negative

³I use the *Q-Learning* algorithm to explain the example. In some reinforcement learning algorithms, rewards are not assigned after each action, but until the very end.

rewards. The applications of these theories are widely adopted in schools and other educational institutions.

Unlike the original studies, the “rewards” in reinforcement learning do not imply any pleasure, hedonic impact, or other psychological stimuli. As we have seen, the reward is a numerical value that does not have any psychological or neurological impact on the agent as is the case with human or animal subjects. The algorithm is therefore out of the realm of the original psychology theory. However, the definitions and the operation of the algorithm produce a narrative that can be perfectly mapped onto the psychology theories. This strategy implicitly helps to validate the plausibility of reinforcement learning as a research area due to the success of its psychological counterparts. More importantly, the narratives of an agent “learning” in the same way that human beings do also contribute to the construction of system intentionality.

System Architecture Level System architecture is another location where intentional narrative can take place. In the late 1980s, the neural network architecture (Figure 15) achieved wide success in various domains, such as computer-controlled autonomous vehicles. Part of the reason for the neural network’s success was not that its performance was statistically better than other machine learning architectures, as many would discover years later. Rather, the hype was built on the narrative that these architectures were based on biological neural structures that humans have, even though, except for the terminology, they only loosely resembled each other. A typical neural network architecture is shown in Figure 15, where the inputs were the stimuli and each of the dots was called a “perceptron.” Other than its original creators who saw neural as a logic model [139], most AI researchers during the neural network boom did not intend to model biological neurons in any rigorous way; what they cared about was solving technical problems with these systems. However, they nevertheless participated in and contributed to the discursive machine of the algorithm. For

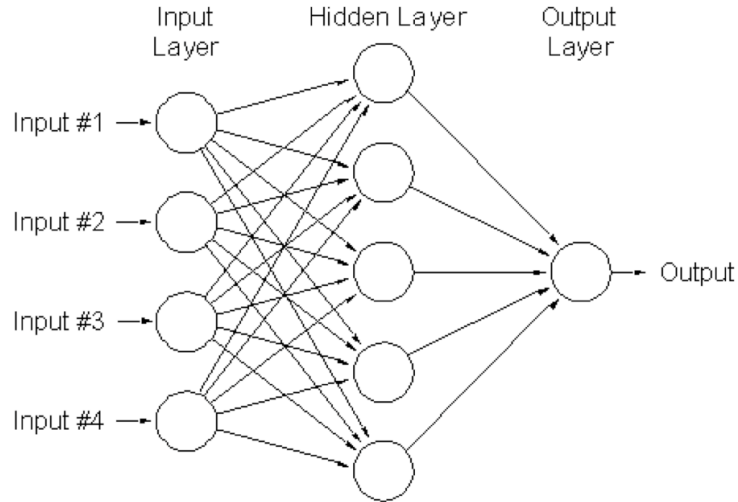


Figure 15: System Diagram of an Artificial Neural Network

obvious reasons, the algorithm’s more descriptive name — the non-linear statistical data modeling algorithm — was hardly ever used either in the community or in mass media. After all, what can be more effective than to say that the computer behind the wheel of the autonomous car is already equipped with neurons and perceptrons to start with?

The boom of neural network research did not last very long, and the neural network paradigm is no longer a key component in mainstream AI research. Interestingly, researchers discovered that, among other issues, the pure numeric results of the neural network architecture do not support sufficient readings and interpretation of its operation. When an AI practitioner works on a planning system, for instance, the actions selected (or not selected) by the system to achieve a certain goal under the given conditions easily form a largely coherent narrative of the system’s operation. This narrative is crucial for the researcher to make sense of the operation of the system. However, a neural network produces only a numerical sequence representing the “weights” at each connection. Debugging any neural network is notoriously difficult because these numbers and the final output of the system cannot be easily connected with an easily understandable narrative.

3.2.4 The Presentation Level

When David Cope released *EMI*'s first CD album *Bach by Design* in 1994, the music was both composed and synthesized by the computer in the tone of a digital piano. By the time the second album *Classical Music Composed by computer* was published three years later, human musicians were hired to interpret and perform the pieces composed by *EMI*. In the following 2000 album *Virtual Mozart*, only historical musical instruments were used in the human performances of the computer generated music.

During the same period of time, the recognition of *EMI* also went through several stages. Although many people, including Cope himself, criticized the completely generated music in the first album as dull and “woody,” the music performed by humans in the second album was said to be lively. *Virtual Mozart* was so well-received that Cope claimed in an interview [99] that “[t]here’s no expert in the world who could, without knowing its source, say for certain that is not Mozart.” This success is due not just to *EMI*'s growing effectiveness, which allows it to compose music “true” to the spirit of the Austrian composer. Cope’s efforts to deploy presentational means also contribute to the aura of the music. As a reviewer of the album describes:

The excellent performance certainly added to the “originality” of the compositions: Only historically authentic instruments were used. The orchestra, directed by the UC Santa Cruz professor Nicole A. Paiement, consisted of two flutes, two oboes, two clarinets, two bassoons, timpani, each two violin I and II, two violas, two celli, and one bass. Even the piano (in the concerto), played by UC Santa Cruz professor Linda Burman-Hall, was a historical instrument from ca. 1800. However, the limitations in the availability of historical instruments were responsible for a re-orchestration: the incorporation of the brass parts into those for the woodwinds.

Even though the AI analysis schema does not address the presentation and interface of intentional systems, the example of *EMI* is a testimony to their importance. Some readers may argue that *EMI* is a special case of an AI-based artistic project and that purely scientific AI research projects manifest their effectiveness through measurable performance, not through presentations. This romanticized notion of scientific practice that transcends human influence, again, has forgotten that AI’s holy grail of the Turing test is itself subjective and interpretive. The following example shows how the discursive machine is present in so called “real” AI projects.

Among the various successors of Weizenbaum’s *ELIZA*, *A.L.I.C.E.* (Artificial Linguistic Internet Computer Entity) is one of the most successful programs. Having won various awards for chatterbots — the *ELIZA*-like conversational computer agents, *A.L.I.C.E.* fits into the “core” AI research agenda of natural language processing. In one of the most famous transcripts of *A.L.I.C.E.*’s conversation with a human, the program is asked to introduce itself. Its reply is “I am an artificial linguistic entity. I was activated at Bethlehem, Pennsylvania, on November 23, 1995. My instructor was Dr. Richard S. Wallace. He taught me to sing a song. Would you like me to sing it for you?” When the human agrees, *A.L.I.C.E.* “sings” the lyrics of “Daisy Bell.” Here, *A.L.I.C.E.*’s creator clearly references the legendary AI system *HAL 9000* in Stanley Kubrick’s film *2001: A Space Odyssey* [114]. “Daisy Bell” was the song that *HAL* sang when it tried to beg the protagonist not to destroy it, one of the most human-like moments of the system in the movie. In the current version of its official website, Wallace also includes an animated anthropomorphized virtual agent (Figure 16). Separately developed by another company, the agent mimics human interaction through its gaze and, in some versions, voice with synchronized lip movements. Although neither the cultural reference nor the anthropomorphized representation directly influence the operation of the natural language processing algorithm, they are part of the discursive machine of the whole system and contribute to the system’s

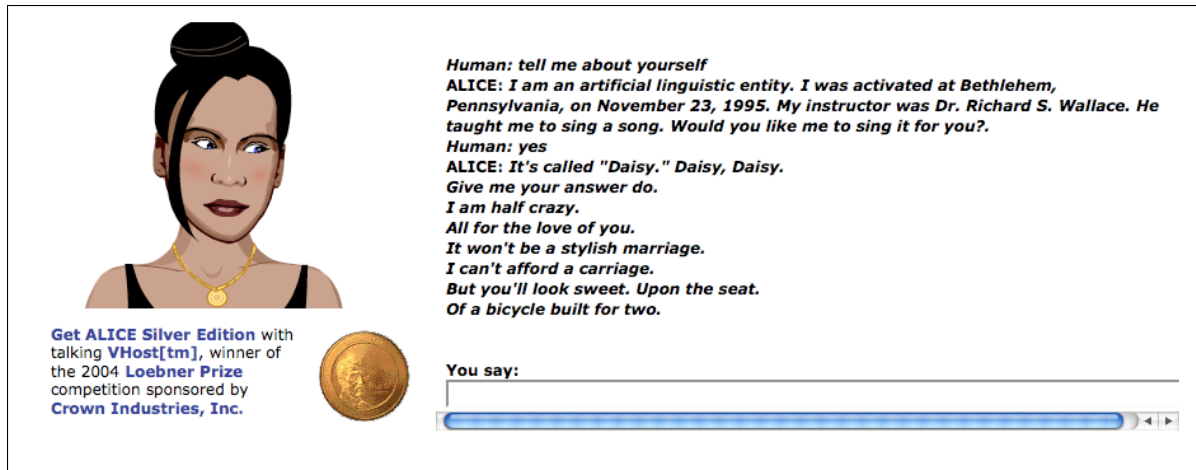


Figure 16: A Partial Screenshot of the A.L.I.C.E Artificial Intelligence Foundation's Website⁵

perception as intentional.

3.2.5 Constructing the Literature

In his book *Science in Action*[122], Latour depicts the techno-science sphere as a two-faced Janus — one for the bracketed and black-boxed *ready made science* that has been accepted as “facts,” and the other *science in the making* as open controversies. Inside the black box of the glorified ready made science innovations in textbooks, Latour demonstrates, are the messiness and contingency of science in the making. Contradicting the view that science and technology are beyond the reach of social and cultural influences, he traces the process of how science is “made.” Based on his ethnographic studies, Latour observes the importance of scientific literature (e.g., publications) that allows scientists to connect both to the scientific community and to other social factors that influence science in the making. He asserts that, “the more technical and specialised a literature is, the more ‘social’ it becomes, since the *number of associations* necessary to drive readers out and force them into accepting a claim as a fact increase” (pp.62). Therefore, in addition to considering actual computer programs and their presentation, an analysis of intentional systems should

include the authors' publications, presentations, interviews, manuals and other interpersonal communication about the system. Such narrative output situates the system in AI research communities and frames users' interpretations and, therefore, must be considered as part of the intentional system.

Technical literature, one of the major venues through which AI practitioners discuss their systems, provides a rich resource for further analyzing the discursive strategies used in the construction of system intentionality. I have conducted an empirical study of this technical literature, paying specific attention to the authors' use of metaphors, blends, and other strategies. The detailed methods and results are discussed at length in Chapter 4.

3.2.6 User's Hermeneutic Reading

Finally, users bring their own experiences and social and cultural backgrounds when they interact with systems in order to appropriate their meanings. In information studies, for instance, researchers have conducted ethnographic studies of how users hermeneutically read quantitative data provided by information systems and how they contextualize these "cold and objective categories and numbers" with the real-life situations. From observing his subjects, Boland [15] draws our attention to the active role users play in appropriating meanings of "text," which he refers to the data provided by the information system used by these subjects. He describes,

These users create a primary context in order to bring the actors and the situation to life, they then proceed to invent alternative contexts and hypothetical histories in order to engage in dialogue with other, possible readings of the text.

A user's reading of an intentional system is based on the system's functionality and her direct interaction with the system. It is also affected by any popular media coverage of the system. The user's expectation and mental model [161] of the system

is influenced by her own personal experience as well as the discourses in society. In this process, the system, intentionality narrated by the authors through various discursive strategies, is interpreted by the users who interact with the system. Although users' interpretations are an indispensable component of the AI hermeneutic network, an extensive user study of their hermeneutic interpretations falls outside of the scope of this dissertation and is a subject for future work.

3.3 A Network Model

So far, we have separately discussed the discursive machine used by system authors to narrate system intentionality and users' hermeneutic interpretations. However, we have to differentiate this framework from the traditional view of text, in which the readers retrieve the meaning encoded by the author. The authors and users of intentional systems do not exist independent of each other, nor are they disconnected from the society. I intend to complicate the model raised at the beginning of this chapter and stress the network nature of the framework.

First, the clear distinction between the user and the author needs to be reconsidered. In the AI community, users and authors are often the same group of people. Many AI systems are meant to be used only by the AI researchers and their graduate students who developed them. In other cases, they may be used by other researchers in the community, whose work (e.g., AI systems and publications) will mostly likely be read by the original system authors.

Second, even if the user is a "layperson," her interaction with and reading of the system and the literature she generates about the system (e.g., blogs) contribute to society's collective experience with intentional systems and a cultural consensus of what constitutes intentionality and intelligence. The moving frontier of AI not only has an impact on AI practitioners, but also speaks to society's ever-changing perception of what is considered intentional. This shift can be manifested through

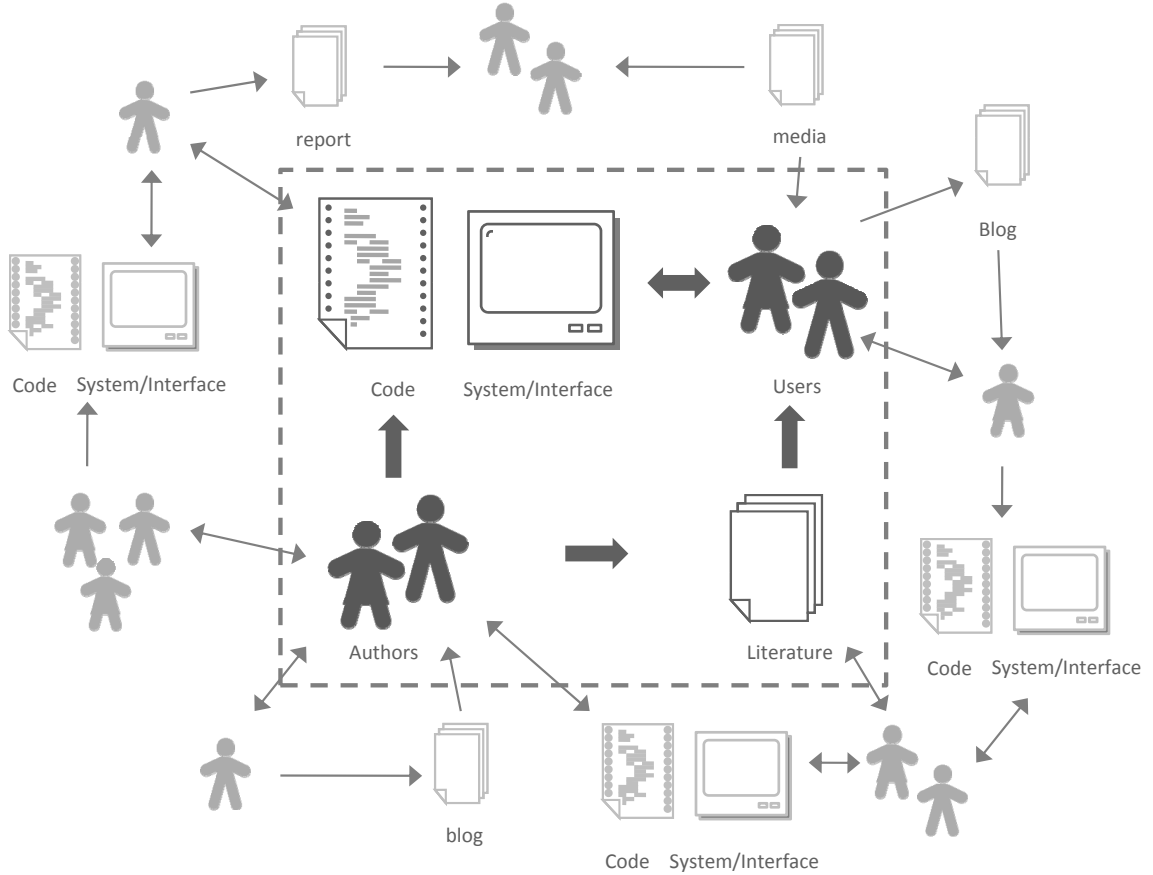


Figure 17: A Network Model of the AI Hermeneutic Network

the vernacular meaning of the intentional vocabulary, and it influences the practice of AI through the duality of these terms.

Third, both previous points have implied a meaning exchange with the authors and users of a specific system to a larger society. Each of the elements in the previous diagram (Figure 13) is connected to a network of other elements. For instance, the development of an intentional system and its rhetorical machine is connected with other similar or related systems. The close reference between *A.L.I.C.E.* and *HAL 9000* is a good example. Similar influences exist in the literature. As the next chapter will show, many authors use intentional vocabulary in accordance with the literature of related works.

Finally, as actor-network theory (Section 2.2.3.2) has shown, other elements should

also be taken into consideration. Government funding agencies, media corporations, public relations managers for the institution where the system was built, an economic crisis, a technological breakthrough, etc. all belong to the AI hermeneutic network.

Although the users' interpretations and temporary suspensions of disbelief are crucial to the formation of system intentionality, this chapter has attempted to reveal the discursive strategies of system intentionality in the technical practice of AI. The construction of intentional systems is often concealed as technical, and its social underpinnings have been generally overlooked. AI systems, as well as many other techno-science practices, are built on the assumption that nature is separate from the society and the human subject, as part of what Latour calls the "purification" process of Modernism [123]. Techno-science, whether romanticized as a savior or demonized as a threat, is commonly deprived of the social context of its construction and seen as answering only to "nature."

As a criticism of this view, this chapter has introduced the new framework of AI hermeneutic network, built on theorizing intentional systems as texts. This framework analyzed intentional systems as social and cultural products, as opposed to construing them as the domain of purely technical practices. It has acknowledged the social and discursive aspects of system intentionality as a result of a hermeneutic communication process that involves both authors' narrations and users' interpretations. This framework opens up the possibilities to expressively manipulate system intentionality in the design of intentional systems. Chapter 5 proposes a new design principle, "scale of intentionality," to demonstrate the new design strategies informed by this theoretical framework.

CHAPTER IV

THE SECRET OF HAPPINESS: A CLOSE READING OF THE *COPYCAT* SYSTEM

... a small renegade codelet, totally unaware of the global momentum, has
had some good luck.... (pp.255)

— Hofstadter, D., & Mitchell, M. “The Copycat Project” [98], 1995

The previous chapter introduced the AI hermeneutic network, which underscores the authors’ discursive narrations and users’ hermeneutic interpretations of system intentionality. This chapter intends to demonstrate the ramifications of this framework, particularly those of the authors’ discursive narrations, through the close reading of an existing intentional system — Douglas Hofstadter and Melanie Mitchell’s analogy-making system, *Copycat*, developed during the early 1990s.

The system intentionality of *Copycat* is manifested through the “insightful” and “creative” analogies the system can come up with as the result of the “fluidity” of its underlying algorithmic operation and knowledge representation. In this chapter, I examine the ways in which authors construct such system intentionality, both technically and discursively. Primary focus here is on the technical literature, a relatively unexplored area for theorists interested in AI, which offers a rich collection of first-hand material of source/pseudo code, architecture diagrams, tables, experiment results, metaphors, and analogies, etc. As we will see from the rest of the chapter, the intertwined technical descriptions and discursive narrations in these articles provide us with a key to understand how system authors navigate between formal and

vernacular narrations to construct system intentionality.

By revealing the discursive strategies embedded in the literature of *Copycat*, my goal is not to attack the “technical soundness” of the system. Its technical contribution and limitation of the project are out of the scope of this dissertation, and have been discussed elsewhere [57, 131, 64, 63]. Instead, my goal of this study is to demonstrate the power and utility of the AI hermeneutic network, in particular, the authors’ narration as a constitutive component of AI and intentional systems, a component which has been ignored by most AI practitioners. As argued in the previous chapter, system authors’ discursive strategies are a constitutive component of all AI and intentional systems, in addition to being part of the systems’ technical operations. Revealing these strategies and making explicit other non-technical factors at work (e.g., social context and authors’ worldviews) contributes to an interpretive and humanistic understanding of system intentionality and AI. The empirical analysis of *Copycat*, as one instance of the many intentional systems, provides a means to understand the narration of system intentionality through a concrete, real system.

Informed by classical hermeneutics, this chapter first introduces an interpretive method to analyze the technical literature, considering the context where the system was developed and the world views of the authors (Section 4.1). In order to gain insight into the latter two elements, I have introduced secondary material (e.g., authors’ interviews with mass media, and their non-technical publications) as a reference. After the main corpus of the study (i.e. technical literature about the system) is introduced, potential limitations of the methods are discussed in Section 4.2. A detailed analysis of *Copycat* is organized according to the method introduced in Section 4.1. First, Section 4.3 discusses the social, technical, and cultural contexts in which *Copycat* was built and their potential influences on the development and narration of the system. Section 4.4 consists of a close reading of the technical literature. As a major step in this analysis, this section demonstrates a discovery, from empirical

data, that intentional vocabulary functions as the connecting point between the technical and intentional narrations, two parallel languages, both present in the corpus. Finally, *Copycat*'s authors philosophical view of AI is analyzed through secondary references in order to provide further understanding of the close reading.

4.1 Method

The method used in the following analysis is informed by classical hermeneutics. As discussed in Section 2.3.1, the long-standing hermeneutic tradition (especially classical hermeneutics) places its emphasis on understanding text according to what its *author intended*. While classical hermeneutics does not acknowledge the plurality of meaning, its focus on authors fits our purpose of unpacking the discursive strategies system authors use in technical literature to narrate system intentionality. After providing a brief account of methods used in one school of classical hermeneutics, this section extends existing hermeneutic methods and describes an integrated method to study the construction of system intentionality in technical literature.

4.1.1 Principles of Hermeneutics

Built on the classical hermeneutic theories, modern day biblical hermeneutics continues to explore the art and science of interpretation. As contemporary hermeneutics scholar Henry Virkler claims, “hermeneutics is needed ... because of the historical, cultural, linguistic, and philosophical gaps that block a spontaneous, accurate understanding of God’s Word” [208]. Similarly, the core idea of hermeneutics can be used to connect the authors and users/analysts of intentional systems, for a similar gap exists between them. Virkler further summarizes hermeneutic principles into five steps. The first three steps constitute general hermeneutics — the study of rules that can be applied for interpreting the entire Bible — and are discussed below. The last two steps are special hermeneutics methods for specific genres, and will not be included in our discussion.

In the first step, the historical-cultural analysis and contextual analysis, a hermeneutic interpreter gains knowledge of the historical and cultural context under which the author wrote the text, as the means to decode the author's "allusions, references, and purpose" of the writing. Contextual analysis, on the other hand, considers the relationship of the specific text and the entire body of texts produced by this author in order to gain a holistic perspective of the author. At this stage, the interpreter's task is to familiarize herself with the historical milieu in which the author wrote the text, to observe the *purpose* of the text based on the author's writing patterns, and to identify the immediate context of the passage under study.

After familiarizing herself with the context under which the specific text was produced, the interpreter is ready to work on the text itself in the second step — the lexical-syntactical analysis. Virkle argues that this step does not encourage blind literalism; it instead differentiates how the author intends the text to be understood, whether literally, symbolically, or figuratively, and informs the interpretation accordingly. Carefully going through the text, the interpreter studies the definitions of words (lexicology) and their relationship to one another (syntax), in order to achieve a concrete and accurate understanding of the intended meaning.

The third major step in the biblical hermeneutical reading attempts to answer the question of how the passage under consideration fits into the total pattern of God's revelation. It assesses the level of theological understanding when the passage was written so that the interpreter can ascertain the text's intended meaning for its *original* readers. Finally, additional knowledge about the topic based on later biblical text should also be acknowledged.

4.1.2 An Integrated Method

The essence of the general hermeneutics methods is to restore authorial intent by carefully studying the context within which a text was created, analyzing the text

itself, and putting the text in the perspective of the author’s ideological world view. This method offers a promising starting point to study a system author’s discursive strategies for creating system intentionality, because the steps to restore an author’s intended meanings can also be used to reveal the strategies the author uses. The rest of this section, based in my theoretical framework, introduces a revised method that integrates hermeneutics, actor-network theory, cognitive semantics, and the philosophy of mind.

Technical-Social-Cultural Analysis Although the study of context can be extended to a much broader set of inquiries, this study limits its scope mainly to the “actants” listed in Latour’s work on science in the making [122]. More specifically, for the particular system under study, I attempt to identify the lab environment where it was created, its funding sources, and the connections between its author and other AI researchers. Also taken into account is the historical development of AI at the time the system was created.

Content Analysis The bulk of the analysis is a close reading of the technical literature about the system. The text of each article in the main corpus is carefully analyzed in order to identify all the rhetorical strategies that are used to construct the system as intentional. Once an article is scrutinized, each instance of discursive strategies is entered into a spreadsheet as a row, with information on the particular intentional vocabulary, the whole sentence where it is used, the article in which this instance appears, and the corresponding page number. Then each row in the spreadsheet is carefully scrutinized again to identify the patterns and trends of these instances of discursive strategies. This spreadsheet then becomes the data for further analytical and statistical examination. The goal of this content analysis is to identify intentional vocabulary and other rhetorical strategies that are deployed —consciously or not— to narrate system intentionality in the presumably “objective” and “neutral”

AI technical literature.

The coding scheme used to identify intentional vocabulary notes any non-trivial token that implies human-like desire, belief, or other intentional mental states, or any token that refers directly to human cognitive faculties. For instance, the sentence “the system itself controls the degree to which it is *willing* to take risks” uses an intentional vocabulary, whereas “SME first *uses* a set of ‘match rules’...” is not considered one because “use” is commonly used in inanimate artifacts. “*Copycat’s Long term memory*” is considered intentional narration, for it directly assigns human faculties to the machine. Of course, sometimes the difference between the intentional and non-intentional vocabulary is a fine line. Although certain terms seem more intentional than others, the prototype model of intentional terms poses a potential limitation on the coding scheme because of the lack of a clear threshold to differentiate them.

Ideology Analysis A study of an author’s rhetorical strategies should also take into consideration her beliefs and values regarding the content she creates. In the context of intentional systems, this means to explicitly identify the author’s position on AI in relation to real human intelligence and intentionality. Is the author a firm supporter of the Strong AI hypothesis? Does she believe that the system she has built leads to achieving machine intelligence? These are important questions in that they help us connect the discursive strategies identified from the previous step to the author’s goal.

This section has introduced a method that is informed by the classical hermeneutic methodology. In the rest of this chapter, this method is used to conduct a close reading of *Copycat*, an analogy-making system, with special attention paid to its authors’ discursive strategies in the technical literature produced about the system.

4.2 *The Corpus and Potential Limitations*

The *Copycat* project was developed by Douglas R. Hofstadter, well-known for his 1980 Pulitzer-winning book *Gödel, Escher, Bach* [93], and his student Melanie Mitchell. It is chosen for close reading for several reasons. First, it is positioned as both an AI and cognitive science project that pushes its authors both to reveal technical advancements and to connect the computational operation to the human mental faculty of analogy. Second, the amount of technical literature generated about *Copycat* offers a rich, yet manageable set of resources for our discussion. The size of the computational analogy community is also well constrained compared to the size of the planning or machine learning communities. Third, the subject of computational analogy is closely related to the cognitively-oriented conceptual blending theories and models used to produce GRIOT and hence *Memory, Reverie Machine* (Chapter 6).

The primary corpus chosen to conduct a close reading of the *Copycat* system comes from the related technical papers and book chapters about the system, which appeared between 1984 and 1995. Among them, four major publications [94, 97, 98, 147] devoted exclusively to this project are chosen as the primary material for this study. (The 1994 article [97] is a combination of the material from the two later ones [98, 147].) These articles were published in 1984, 1994, 1995, and 1995 respectively by Hofstadter and Mitchell, incorporating over 200 pages of technical descriptions of *Copycat*, comparing it to other computational analogy systems, particularly the *Structure-Mapping Engine (SME)* [50] and *ACME* [101]).

In addition, other related literature, including *Copycat's* main author's (Hofstadter) interviews in the mass media, non-technical articles, websites are also studied, particularly as with regard to the social contexts and his ideological beliefs regarding issues related to intentional systems. Admittedly, several limitations exist in terms of the corpus. First, certain articles, which may deploy discursive strategies differently, could be left out since the corpus does not exhaustively include all publications on

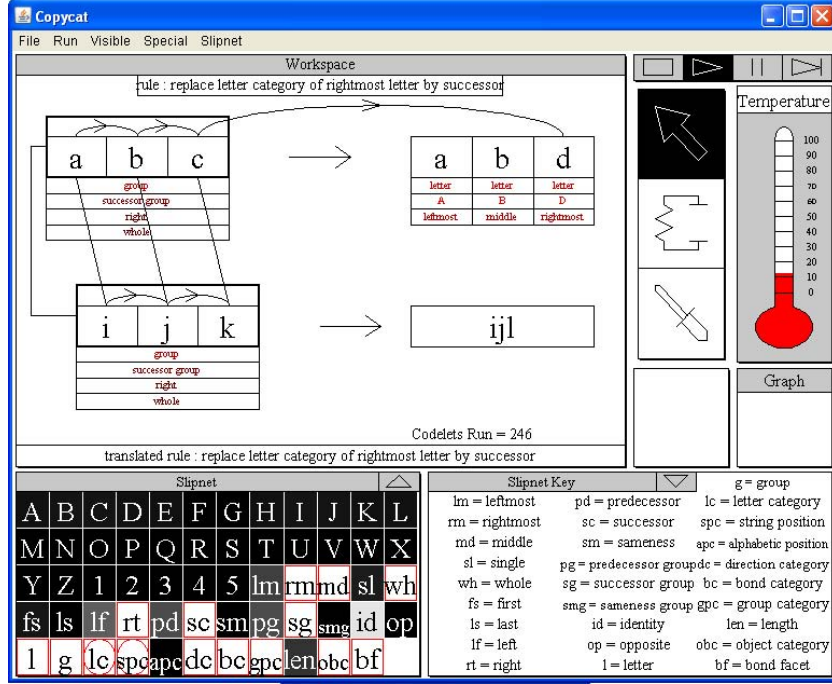


Figure 18: Screenshot of a Java Re-implementation of *Copycat* by Scott Bolland¹

Copycat. Second, other AI practitioners' discursive strategies may be very different from those of *Copycat*'s authors. Although these issues may potentially undermine the generalizability of the results of this analysis, the aim of this analysis is not to propose a generic pattern that fits all AI practitioners. In fact, part of the function of the context and ideological analyses is to take into account the individuality of each system's author.

4.3 The Technical-Social-Cultural Context of Copycat

The development of the *Copycat* project coincides with the two AI winters — a term coined as an analogy to the idea of nuclear winter. During the two AI winters, both government/military funding and public interest for AI dropped substantially due to the frustrations caused by the unrealistic expectations and over-promising of the field. This chain reaction began with pessimism in the AI community, followed by pessimism in the press, followed by a severe cutback in funding, followed by the end of serious research [32, pp.203]. Some of the major AI developments around the time

when *Copycat* was created are listed in Table 6.

After the failure of machine translation and the abandonment of connectionism in the late 1960s and early 1970s, DARPA canceled many of its AI projects, including a three million-dollar a year grant for Carnegie Mellon University’s Speech Understanding Research. J. C. R. Licklider’s “funding people, not projects” policy in the 1960s was replaced by a more pragmatic and applied approach; undirected research with no direct applications for the military had a hard time receiving DARPA money. Things got even worse when Sir James Lighthill released a report for the British Parliament in 1973, stating that nearly no AI projects would generate truly usable technology in the foreseeable future. Soon enough, government funding for AI became hard to find and hence the first AI winter (1974 - 1980).

Thanks to the commercial successes of expert systems in the beginning of the 1980s, the AI community slowly recovered from the aftermath of the first AI winter. Many companies around the world quickly adopted these systems, which simulated the reasoning process of human experts in solving problems such as medical diagnoses (*MYCIN* [197]) and the evaluation of mineral potential (*PROSPECTOR* [44]). However, as these companies would soon find out, these domain-specific programs were too expensive to maintain. In addition, AI faced other major setbacks, such as the fall of the LISP machine markets and the failure of Japan’s ambitious Fifth Generation Computer (logic programming) project. The result was that the field entered its second winter (1987 - 1993).

Although never mentioned explicitly in the technical literature itself, the context of the two AI winters had a very direct impact on the *Copycat* project. Technically, the two AI winters marked a paradigm shift of the community from symbolic AI (GOFAI) to various different approaches and eventually to interactionist AI. Developed during this transitional period, *Copycat* embodies techniques, values, and rhetoric from multiple approaches to AI and cognition.

Table 6: Timeline of Major AI Developments between 1980 and 1999

Year	Major AI Developments
1980	The First AAAI conference. First expert system shells and commercial applications developed.
1981	Danny Hillis’ the connection machine, a massively parallel architecture, brought new power to AI and computation in general.
1985	Harold Cohen’s AARON was demonstrated at the IJCAI conference. Neural Networks become widely used with the Backpropagation algorithm.
1986	SME, a renown analogy system, was published by Falkenhainer, Forbus and Gentner.
1987	Marvin Minsky published <i>The Society of Mind</i> , a theoretical description of the mind as a collection of cooperating agents. Rodney Brooks introduced the subsumption architecture and behavior-based robotics as a more minimalist modular model of natural intelligence; Nouvelle AI.
1994	With passengers onboard, the twin robot cars (VaMP and VITA-2) of Ernst Dickmanns and Daimler-Benz drove more than one thousand kilometers on a Paris three-lane highway in standard heavy traffic at speeds up to 130 km/h.
late 1990s	The first webcrawlers appear. Along with other AI-based information-extraction programs, they become critical to widespread use of the World Wide Web.

It is true that *Copycat*’s domain and high-level goal of symbolic analogy between short strings of letters lies at the heart of the symbolic AI paradigm. Even its concept net is reminiscent of classic symbolic systems such as Douglas Lenat’s *CyC* [125] and Minsky’s concept of frame systems [143]. Yet, in comparison with its contemporary competitor *Structural Mapping Engine* (*SME*) [50] by Falkenhainer, Forbus, and Gentner, *Copycat* is not heavily invested in the syntactic rules of analogy, regardless of the content of the problem. Even its concept net has a flavor of the connectionist approach by introducing variable weights to the links between concepts. As other AI researchers [131] have commented, “[c]ompared to Gentner’s structure-mapping theory of analogy emphasizes formal, shared syntactic relations between concepts. In

contrast, Hofstadter and Mitchell’s ‘slipnets’ [(*Copycat*)] project emphasizes semantic similarities and employs connectionist notions of conceptual distance and activation to make analogy more dynamic and cognitively plausible.”

But overall, the architecture and vocabulary used in *Copycat* speak strongly to a dynamical systems approach. Unlike symbolic AI (Section 2.1.1), the dynamical systems approach does not emphasize the syntactic manipulation of symbolic representations of a problem. Instead, its primary focus is how a system changes its states over time. Randall Beer describes:

the explanatory focus [of a dynamical model] is on the structure of the space of possible trajectories and the internal and external forces that shape the particular trajectory that unfolds over time, rather than on the physical nature of the underlying mechanisms that instantiate this dynamics. On this view, inputs do not uniquely specify an internal state that describes some external state of affairs. Rather, they serve as a source of perturbations to the system’s intrinsic dynamics. [11]

The dynamical systems approach’s emphasis on the evolution over time and state, “self-organization and the emergence of structure,” and “multiple simultaneous interactions” among different parts of the system [47] lends itself very well to Hofstadter’s notion of fluid analogy. For instance, the system contains a pool of codelets (small pieces of code that function like autonomous agents) including both top-down and bottom-up functionalities, working simultaneously to provide control and to explore new potential discoveries. The stochastic behaviors of the codelets over time, according to its authors, provide the system with the necessary flexibility to discover different kinds of analogies.

Some readers may find some parallelism between the *Copycat* architecture and Marvin Minsky’s book *The Society of Mind* [144]. To a certain extent, both works are about how intelligent behaviors can emerge from a large group of simple and

unintelligent elements. In fact, Hofstadter spent a sabbatical year in 1983 at MIT's Artificial Intelligence Lab with the support of Minsky. In this year, Hofstadter wrote the first full article [94] on *Copycat*, published the following year as a technical report of the lab. In this 1984 article, the major ideas of *Copycat* were already proposed. It was also during his stay at MIT that Hofstadter met Melanie Mitchell, who subsequently did her Ph.D. on *Copycat* at Indiana University with Hofstadter.

Copycat's main sources of funding came from the National Science Foundation, Indiana University, University of Michigan, and several personal funding foundations in the early days of the project. The lack of substantial DARPA funding meant that Hofstadter had to find ways to fund his project, but also meant that he had a relatively large amount of intellectual freedom to pursue his own interest without having to tie it to military or other commercial applications. In the Prologue of his second book, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, expressed his frustration with the field of AI at that time:

For me and probably for a good many other people, the term [artificial intelligence] conjured up an exciting image — that of questing after the deepest secrets of the human mind and expressing them as pure, abstract patterns. In the early 1980's, however, that term, as words are wont to do, gradually started changing connotations, and began to exude the flavor of commercial applications and expert systems, as opposed to basic scientific research about the nature of thinking and being conscious. As a result I came to feel much less comfortable saying or writing "AI." Luckily, a new term was just then coming into currency — "cognitive science" — and I started to favor that way of describing my research interests, since it clearly stresses the idea of fidelity to what actually goes on in the human mind/brain, as well as the pure-science nature of the endeavor. [95]

This somewhat ambivalent position of *Copycat* as half-AI and half-cognitive science research had an impact on the discursive strategies the authors use in their technical publication, as we will see in the next section.

4.4 *Content Analysis of the Technical Literature*

After a brief account of the technical, social and cultural contexts in which *Copycat* was developed, this section zooms in on the analysis of the main corpus — technical literature of *Copycat*. Once again, teasing out the discursive elements in the system may seem from time to time a hostile gesture towards the system. However, as explained earlier, this is *not* the goal of this study. In fact, the influence of this well-known system in the history of AI is itself a proof of the significance and contribution of the work as well as its authors. In fact, the potential reaction to seeing this study as an “attack” illustrates the pervasiveness of the techno-centric view in the AI community — if the intentionality is not from the technical operation of the system, it is not valid AI research. My goal then is to reveal the existence of the discursive component in what has been regarded as purely technical systems and to argue that they are as constitutive as technical operation to these systems.

In this section, Section 4.4.1 first describes *Copycat*’s operation in two different kinds of languages, one intentional and the other technical, in order to demonstrate the existence of both in the main corpus. The next section (4.4.2) demonstrates that the intentional vocabulary offers system authors a powerful link that connects the two languages that AI authors have to master simultaneously. Then, in Section 4.4.3. patterns of system authors’ navigation between the intentional and technical languages are discussed.

4.4.1 The Two Aspects of *Copycat*

Before diving in-depth into corpus, it is necessary to familiarize the readers with the high-level account for the operation and architecture of the system. Although the

corpus offers mixed descriptions of the system, this section attempts to introduce *Copycat* from two perspectives. The first description is constructed with intentional vocabulary, drawn mainly from the corpus. The second version is mainly a technical one with very few intentional terms, based on the source code and my understanding of the system as a computer scientist.

The reason for juxtaposing the two different descriptions is to highlight the discursive machine’s impact on the perception of system intentionality. What is striking is not merely that the two descriptions seem to be about different systems, but that they seem to depict different *kinds* of systems. With the formal and technical terms, most people would not expect that the second version describes an AI system.

4.4.1.1 *A Fluid Analogy Maker*

The *Copycat* project is an attempt to model human “mental fluidity” in analogy making. Its authors claim that the program is capable of making “insightful” analogies in a small, restricted domain, namely alphabetic sequences. For instance, when *Copycat* is shown the following “problem”: **abc** \rightarrow **abd**, **pqr** \rightarrow ?, *Copycat*’s answer will likely be **pqs**. This domain is not as simple as it might seem, as certain analogies are challenging even for humans. For instance, faced with a new problem **abc** \rightarrow **abd**, **pqrrrr** \rightarrow ?, *Copycat* finds the answer **pqrrrrr** by mapping *alphabetic position* to *group size*. When the result is shown to humans, many find it “insightful.” *Copycat*’s functioning is non-deterministic; in different runs, it may find different solutions to the same problem... In *Copycat*’s domain of operation, many solutions are valid for each problem, depending on how the subject “interprets” the problem. Solutions that are more “obvious” are found more often, whereas other solutions are found with lower frequency.

In order to model human “conceptual fluidity” in perception and analogy making, *Copycat*’s main focus is the “slippage” of concepts into one another. For instance,

in the previous example, each letter was seen as its numeric position in the alphabet so when the **c** was changed to a **d**, *Copycat* understands that the group of 3 **rs** has to be changed to a group of 4 **rs**. In order to achieve this, the representation of the system’s perception is constructed dynamically. When the system receives **abc** as the input, *Copycat*’s perception mechanism constructs a representation containing information such that **abc** is a string composed of three letters or that **b** is the alphabetical successor of **a**. *Copycat* contains a collection of “concepts” (e.g., “successor,” “predecessor,” “length,” “alphabet,” “sameness,” etc.) with which it “perceives” the input. Depending on the problem the system is dealing with, some concepts might “slip” into one another, and *Copycat* might all of a sudden map “alphabetic position” to “group size” (as in the previous example). Finally, the system is also capable of explaining its predictions by presenting to the user which concepts are mapped to which other concepts.

Copycat incorporates three different modules: the “slipnet,” locus of the concepts that the system has access to; the “workspace,” where *Copycat* constructs the representation of the problem it is solving, and where the analogy is made; and finally the “coderack,” where a collection of *codelets* is waiting to be executed. Codelets are small pieces of code that perform tasks such as creating or destroying a new perceptual structure, evaluating how promising a particular structure is, or creating more codelets. Codelets are executed stochastically, and the exhibited behavior of *Copycat* is the result of the execution of hundreds of codelets. As the authors explain, codelets can be seen as the enzymes in biological cells, where each enzyme does only one very small task, but the combination of thousands of them manages to fulfill complex tasks. The individual objects constituting the representation being built in the workspace (corresponding to how *Copycat* is perceiving the input) are assigned an “unhappiness” value (corresponding to how unsatisfied the system is with the structures associated to that object). When the unhappiness decreases, a global measure

of “temperature” decreases. The lower the temperature, the more likely the program will terminate, and generate an analogy based on the current representation in the workspace.

Compared with other models of analogy in AI or cognitive science (e.g., *SME* [50] or *ACME* [101]), *Copycat* has two main distinctive features. First, it models both the perception and mapping stages of analogy (retrieval, perception, mapping and inference are the four well established phases of analogical reasoning [50]). Second, its biologically inspired search scheme based on codelets is claimed to be “psychologically plausible” by its authors, in opposition to more traditional exhaustive search methods used by all other models of analogy.

4.4.1.2 *A Stochastic Local Search Program*

So far, we have seen the intentional narrative that the authors of *Copycat* have constructed. In order to highlight the effect of these discursive strategies, this section undertakes the task of formulating a technical description of the system’s operation, using conventional terms that are formally defined and adopted by the computer science community.

Copycat is a stochastic local search program. It receives three character-based strings (String 1, String 2, and String 3) as input and generates a single output character string. In order to generate the output character string, *Copycat* performs a stochastic local search in a particular search space, optimizing a particular heuristic function. The search space is the space of all possible structures that relate the three input strings together. Each one of these structures is a graph built from a base set of primitive constructs predefined by the author (such as “b is the successor of a”). A particular structure captures the relations among the three input strings and determines the compatibility of the different primitive constructs appearing in a structure. *Copycat* maximizes a heuristic function, that is, the extent to which

the proposed structure captures all the regularities and relations among the three strings. The system may randomly terminate its search at any point in time. The probability of termination is higher if the system has found a structure with a high value based on the heuristic function. Once the search stops, the system generates an output string according to the transformation operations specified by the current structure. This means that the same operations that transform String 1 to String 2 will be applied to String 3 in order to derive the output string. For instance, when *Copycat* receives the following input: **abc**, **abd**, **pqr**, *Copycat*'s answer will be **pqs** with a high probability.

Although the above technical description only seems to loosely relate to the first intentional description, it captures the operation of the system. For instance, in the original description of *Copycat*, the authors state that the system simulates human perception, and that it builds a “point of view” of the input problem. This means that the system builds a structure that relates all the letters in the three input strings. These structures are assigned a “strength”, which is a function of its “happiness,” “conceptual depth,” and “contradiction.” (Two structures are contradictory, when they contain contradictory “slippages,” e.g., in one “successor” slipped to “predecessor” and in the other one it did not.) Depending on the strength of the structures currently present in the workspace, *Copycat*'s “temperature” is higher or lower. As the temperature decreases, the chances of *Copycat* stopping are higher. Thus, the effect is that the system is optimizing its structure strength.

A comparison between the above two aspects is illustrated in Figure 19. I do not wish to imply that one description is more “objective” or “accurate” than the other. As argued from the last chapter, the choice of discursive strategies is an intrinsic part of AI practice. The previous paragraph foreshadows the next section, which demonstrates how the authors of *Copycat* connect the intentional and technical languages by using intentional vocabulary.

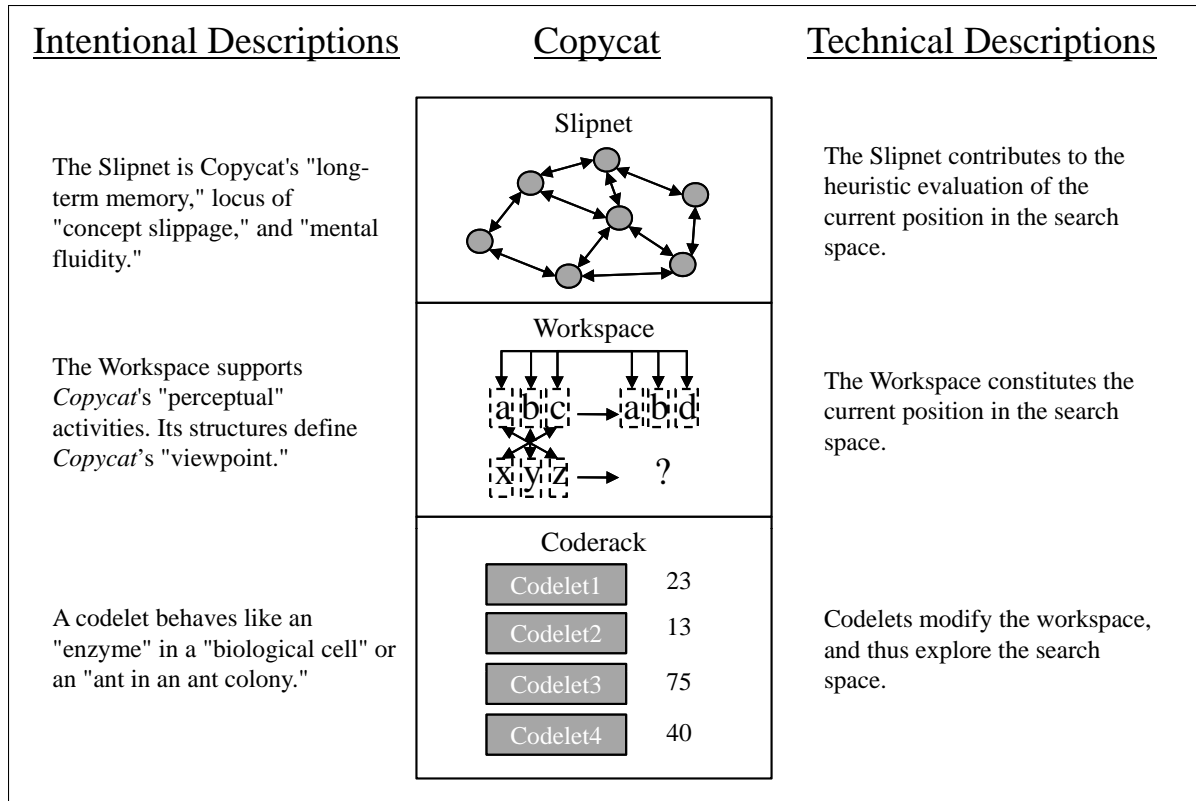


Figure 19: A Comparison between the Intentional and Technical Descriptions of *Copycat*

4.4.2 Intentional Vocabulary: Connecting the Two Languages

This section presents one of the main arguments of this chapter, that intentional vocabulary is essential to the practice of AI in that it provides a connecting point between the technical language and intentionality language, both of which are part of AI literature. First, Section 4.4.2.1 presents the empirical data from the analysis of the corpus and categorizes them into three different uses of intentional vocabulary. Then we closely trace two examples of “*scout* codelets” and the “*happiness* value of codelets” in order to understand how the authors use intentional vocabulary to leverage the technical and discursive requirements.

4.4.2.1 The Use of Intentional Vocabulary

One of the main distinctions of *Copycat* from other existing computational analogy systems, in the eyes of its authors, is that *Copycat* offers a “psychological plausible” model of analogy. Although the authors never pin down the exact meaning of this goal, it is repeated many times in the literature: “we have been influenced more by *psychological* than neurological findings,” “*SME* is an algorithmic but *psychologically implausible* way of finding what the structure-mapping theory ...” or “certainly, the exhaustive search *SME* performs through all consistent mappings is *psychologically implausible*” (emphasis added).

In order to demonstrate the “psychological plausibility,” a close parallel between *Copycat* and human cognitive processes is made through various discursive strategies that the authors deploy. Among others, three strategies used frequently by the authors stand out. First, intentional verbs are heavily used to narrate the operation of the system. Such words as “know,” “resist,” “understand” appear throughout the corpus. More examples include (emphasis added):

Just as the program *knows* the immediate neighbors of every letter in the alphabet, it also *knows* the successors and predecessors of small integers.

Copycat tends to *resist* bringing numbers into the picture, unless there seems to be some compelling reason to do so.

Musing codelets allow several different and rival pathways to be *sniffed* or *checked out*.

Second, certain data structures and functions are named after human cognitive faculties and human mental states. If the previous strategy is concerned with the intentionality of the system, the narrations at this level are intended to draw close connection between *Copycat*’s operation to human cognitive process and lay the groundwork

for the next strategy. For instance, the search space of *Copycat*, a term hardly used in the corpus, is composed of structures called “point of views” (or simply “views”), which specify the ways different “concepts” connecting the three input strings (e.g., **abc**, **abd**, and **pqr**). Similarly, *Copycat* has “long-term memory,” “drive,” “desire,” and “personality:”

It [(the Slipnet)] can be thought of, roughly, as *Copycat*’s *long-term memory*.

[*Copycat*] must reconcile a large number of mutually incompatible local *desires* (the technical term for this is ‘*frustration*’).

... and those data provided some of the most important insights into the program’s “*personality*.”

Last but not least, the system is often narrated in comparison to human and other forms of life (e.g., cells, and ants). Although the content may vary depending on the context, these arguments typically take the following form: A (creative) human faced with situation X will react with action Y, and *Copycat* also performs action Y in this situation X. The purpose of drawing such a comparison is to imply that *Copycat* is similarly creative, intelligent and intentional. An example is:

In particular, people are clearly quicker to recognize two neighboring objects as identical than as being related in some abstract way. Thus the architecture has an intrinsic speed-bias in favor of sameness bonds: it tends to spot them and to construct them more quickly than it spots and constructs bonds representing other kinds of relationships.

The following tables 7, 8 and 9 respectively illustrate examples of each of the three above mentioned discursive strategies appearing in the corpus. More complete data can be found in the Appendix. Readers can get a direct sense of the deployment of

Table 7: Usage of Intentional Vocabulary

Context	Intentional Term	Source
Just as the program <i>knows</i> the immediate neighbors of every letter in the alphabet, it also <i>knows</i> the successors and predecessors of small integers.	know	[98, pp.210]
Thus, copycat is capable of <i>recognizing</i> not only that the structure ...	recognize	[98, pp.210]
Copycat tends to <i>resist</i> bringing numbers into the picture, unless there seems to be some compelling reason to do so.	resist	[98, pp.210]
This is how the system’s actual <i>viewpoint</i> develops with time.	viewpoint	[98, pp.226]
the system loses this naive, <i>open-minded</i> quality; as indeed it ought to and usually ends up being <i>closed-minded</i> ...	open-minded	[98, pp.228]
Thus, the system itself controls the degree to which it is <i>willing</i> to take risks.	willing	[98, pp.229]
... and those data provided some of the most important insights into the program’s <i>personality</i> .	personality	[98, pp.230]
The answer xyz , whose very high temperature of 74 indicates that the program did not “like” it at all, ...	like	[98, pp.246]

authors’ narration from these examples. Grounded in these findings, the next section offers a more in-depth analysis of the trends and patterns of how the authors narrate the system intentionality of *Copycat*.

4.4.2.2 The Scouts

In the Copycat corpus, certain types of codelets called the “scouts” or “musing codelets” receive a lot of attention. According to the articles, their purpose is to look for potential connections and relations (e.g., “sameness”) between objects in the working space for the main system to further scrutinize and evaluate. The image of human scouts is familiar to most people — they are sent ahead of a main force to gather information about the enemy’s location, strength, etc. In the following depictions of the codelets, the authors clearly leverage our mental image of a human scout:

Table 8: Usage of Cognitive Faculties in Naming Functions and Structures

<i>Sentence</i>	Term	Source
It (the Slipnet) can be thought of, roughly, as Copycat’s <i>long-term memory</i>	long-term memory.	[98, pp.211]
The Workplace is the locus of <i>perceptual activity</i> .	perceptual activity	[98, pp.211]
An unhappy object is one that has few or no connections to the rest of the objects in the Worldspace, and that thus seems to cry out for more attention	unhappy	[94, pp.217]
The topology of our Slipnet is defined by certain <i>concepts</i> realized as nodes	concepts	[94, pp.4]

The basic image is that of many “fingers of exploration” simultaneously feeling out various potential pathways at different speeds, thanks to the coexistence of pressures of different strengths. These “fingers of exploration” are tentative probes made by scout codelets

The foggy things are, the more unbiased should be the scouting mission

The mission of such a codelet is to scan the Workspace in search of instances of its spawning concept.

A metaphor for the parallel terraced scan is provided by the image of a vast column of ants marching through a forest, with hordes of small scouts at the head of the column making small random forays in all directions (although exploring some directions more eagerly and deeply than others) and then returning to report; the collective effect of these many “feelers” will then determine the direction to be followed by the column as a whole. This is going on at all moments, of course, so that the column is constantly adjusting its pathway in slight ways.

In addition to the explicit intentional vocabulary (e.g., “explore,” “feel out,” “mission,” “march,” and “report”), the authors use many terms with spatial connotation

Table 9: Comparison of *Copycat* with humans and other forms of life

<i>Sentence</i>	Source
Creative human thinkers manifest an exquisite selectivity of this sort — when they are faced with a novel situation, what bubbles up from their unconscious and pops to mind is typically a small set of concepts that “fit like a glove,” without a host of extraneous and irrelevant concepts being consciously activated or considered.	[98, pp.211]
This image comes essentially from the biological cell; the Workspace corresponds roughly to the cytoplasm of a cell, in which enzymes carrying out diverse tasks all throughout the cell’s cytoplasm are the construction crews, and the structures built up are all sorts of hierarchically-structured biomolecules.	[98, pp.216]
In particular, people are clearly more quickly able to recognize two neighboring objects as identical than as being related in some abstract way. Thus the architecture has an intrinsic speed-bias in favor of sameness bonds: it tends to spot them and to construct them more quickly than it spots and constructs bonds representing other kinds of relationships.	[98, pp.218]

such as “potential pathways,” “forest,” “direction.” Other terms also have strong tie to space: “foggy,” “scan,” “search,” and “exploration.” Embodied cognition theories have stressed the importance of our motor-sensory experience as essential to human intentionality and understanding [88]. Lakoff [118] also illustrates how the fundamental image schema such as “CONTAINER” and metaphors such as “LIFE AS A JOURNEY” are temporally and spatially grounded. Similarly, the spatial aspects in the description of the scout codelets make them easy to relate to and therefore easy to see as intentional.

However, the spatial narration is brought in from the conceptual spaces of the “human scout” and the “ant scout.” As the source code of a particular kind scout codelet shows in Figure 20, the system has no spatial concepts at all. As a close analysis of Melanie Mitchell’s dissertation [146] echoes, these codelets do not “scout” or “explore” or have “missions” or goals. Their operation is to randomly pick one object in the workspace, and a concept in the slipnet, and check whether the selected concept can be applied to the selected object. Moreover, the search space of the scout

```

(define-codelet-procedure* bottom-up-bond-scout
  (lambda ()
    (let* ((from-object (tell *workspace* 'choose-object
                              'get-intra-string-salience))
           (to-object (tell from-object 'choose-neighbor)))
      (if* (not (exists? to-object)) (fizzle))
      (let ((bond-facet (choose-bond-facet from-object to-object)))
        (if* (not (exists? bond-facet)) (fizzle))
        (let* ((from-object-descriptor (tell from-object
                                              'get-descriptor-for bond-facet))
                (to-object-descriptor (tell to-object 'get-descriptor-for
                                              bond-facet))
                (bond-category (get-bond-category from-object-descriptor
                                                  to-object-descriptor)))
          (cond
            ((not (exists? bond-category)) (fizzle))
            ((incompatible-bond-candidates? from-object to-object bond-facet
                                              bond-category)
             (fizzle))
            (else (propose-bond from-object to-object bond-category bond-facet
                                from-object-descriptor to-object-descriptor)))
          )
        )
      )
    )
  )
)

```

Figure 20: Source Code for a Bottom-up Scout Codelet

codelets is very constrained. For instance, the scout codelets look only for relations among contiguous letters in a string, which drastically reduces the search space.

Notice that the operation of codelets is quite different from “hordes of ants exploring a forest.” The discontinuous and random combination of concepts and relations is repurposed and narrated through the continuous spatial exploration of a human/ant scout. It is also worth pointing out the co-existence of multiple narratives inside the code shown in Figure 20. In addition to the intentional vocabulary, chemistry-related (e.g., “bond,” and “fizzle”) and other (e.g., “workspace,” “category,” and “descriptor”) narrations are also present. These multiple and intersecting metaphor systems complement the primary narration of system intentionality, and offer a spectrum of narratives, from the more intentional to the more technical ones. The functioning of this field of narratives, composed of different metaphors and semiotic systems, is a rich area for further exploration.

The intentional vocabulary does not always bring new properties that do not exist in the machine operation to the narration of system intentionality. As we will see from the following example, its more important function is to provide a bridge between the discursive layer and the technical layer of the system.

4.4.2.3 *The Reason of “Happiness”*

Most AI practitioners will be quick to agree that “happiness” is an explicit intentional term. Unlike terms such as “planning” or “learning,” “happiness” does not have a conventionally agreed upon formal meaning that the AI community follows. In fact, its highly subjective and emotional undertone is almost an antithesis of any formal definition based on machine operation. This section provides an example of how the authors of *Copycat* use intentional vocabulary such as “happiness” to connect the intentional narration needed for their research goal and the technical requirement for the systems.

In *Copycat*, the level of “happiness” of each object (e.g., a letter, a number, or a group of letters from the three input strings) in the system’s “work space” is an important factor. It regulates how much attention the system pays to a specific object; the unhappier an object is, the more resources will be given to it. As its authors describe,

Unhappiness is a measure of how integrated the object is with other objects. An unhappy object is one that has few or no connections to the rest of the object in the Worldspace, and that thus seems to cry out for more attention.

The choice of an emotional term with easily understandable meaning is far from an accident. It reinforces the overarching research goals set by *Copycat*’s creators, that is, to model the analogy-making process with psychological plausibility. As the authors position the system in one of the opening paragraphs, “Copycat is a

computer program designed to be able to discover insightful analogies, and to do so in a psychologically realistic way.” One effective strategy to accomplish goal is to connect the system’s operation to with common wisdoms. The authors wrote: “the architecture follows the old motto ‘The squeaky wheel gets the oil’, even if only probabilistically so.” Following this motto, it seems “natural” and “human-like” that the “unhappy” concepts is entitled to more attention.

However, the “unhappiness” measure also serves a technical purpose that is never explicitly mentioned in the corpus. Technically speaking, *Copycat*’s goal is to optimize the overall connection between different objects in the working space. The strength of each structure is computed as an aggregation of the strengths of the individual elements (e.g., bonds) in the structure. An “unhappy” element corresponds to an element with weak structures, whereas a “happy” one has an already formed strong structure that connects it to other elements. In probabilistic terms, working on the weakest point in a structure yields the most chances for improvement; for modifying a strong structure is likely to make it weaker (since the structure is strong), and modifying a weak structure is likely to make it stronger (since it is already weak). Hence, focusing on the “unhappy” objects maximizes the probability of strengthening the current structure.

Figure 21 provides the source code for one particular kind of “happiness.” The implementation shows a rather simple function with little connection to the vernacular meaning of the term. Here, the so-called “intra-string-unhappiness” is updated with a number between the range of 0 and 100, depending on the properties of the connection this element has. As stated earlier, the point of the example is not the disjunction between the intentional narrative and the computational operation. Rather, it illustrates the intentional vocabulary’s pivotal function of connecting the two. Without the intentional narration, the technical machine operation lacks the system intentionality necessary for intentional and AI systems. Without the machine operation, on

```

(record-case (rest msg)
  ...
  (update-intra-string-unhappiness ()
    (set! intra-string-unhappiness
      (cond
        ((tell self 'spans-whole-string?) 0)
        ((exists? enclosing-group)
          (100- (tell enclosing-group 'get-strength)))
        (else
          (let ((bonds (tell self 'get-incident-bonds)))
            (cond
              ((null? bonds) 100)
              ((or (tell self 'leftmost-in-string?)
                    (tell self 'rightmost-in-string?))
                (100- (round (* 1/3 (tell (1st bonds)
                                          'get-strength))))))
              (else
               (100- (round (* 1/6
                             (sum (tell-all bonds
                                     'get-strength))))))))))
      'done)
    ...
  )
)

```

Figure 21: Source Code for One Kind of “Happiness”

the other hand, intentional narrations are vague and hard to believe. The intentional terms such as “happiness” provide a joint so that the discursive and computational operations can cling to each other.

4.4.3 Leveraging the Two Languages of AI

In the previous chapter, we discussed the dual roles of an AI practitioner as both the executor and narrator of her system. On the one hand, the practitioner faithfully follows the established formal conventions and methodology associated with the key AI intentional terms. On the other hand, she leverages the vernacular meaning of these terms to construct the narrations of system intentionality. So far in this chapter, we have seen the existence of the two languages of AI — one intentional and the other technical — linked by certain intentional terms. This section documents patterns of the circumstances in which each language is adopted by the authors of *Copycat*.

An overall pattern of the authors’ strategy is that the intentional narrations are

generally reserved for setting up the goal of the system and describing the system at a high level, whereas more technical descriptions are adopted when *Copycat* is compared with other similar systems. In the opening paragraphs of all four articles, intentional narrations of the system are used extensively to reinforce the connection between *Copycat* and human cognitive processes. For example, in one of the articles, the authors wrote “... the Copycat project is not about simulating analogy-making per se, but about simulating the very crux of human cognition: fluid concepts” [98, pp.208]. Another example is, “Just as the program knows the immediate neighbors of every letter in the alphabet, it also knows the successors and predecessors of small integers” [98, pp.210].

Figure 22 and 23 show the occurrences and density of intentional vocabulary in “The Copycat Project: A Model of Mental Fluidity and Analogy-making” [98], defined as the ratio of the number of intentional vocabulary used to the entire word count of each section. The titles of each section, in sequence, are “Copycat and Mental Fluidity” (Section 1), “The Three Major Components of the Copycat Architecture” (Section 2), “The Emergence of Fluidity in the Copycat Architecture” (Section 3), “The Intimate Relation between Randomness and Fluidity” (Section 4), “Copycat’s Performance: A Forest-level Overview” (Section 5), “Copycat’s Performance: A Tree-level Close-up” (Section 6), and “Conclusion: The Generality of Copycat’s Mechanisms” (Section 7). The figure shows a higher density of intentional vocabulary in Section 1, 3, and 4 when the topic of *Copycat*’s close relationship with mental fluidity is discussed. The sections with lower density (Section 5, 6 and 7) are concerned with topics of system performance, and conclusion.

In contrast, when *Copycat* is compared with other computational analogy systems, very few intentional terms are used. In the article “Perspectives on Copycat: Comparisons with Recent Work” [147], the use of intentional vocabulary is 0.76 / page, compared to 1.5 / page in another article [98] listed in the same book. In other

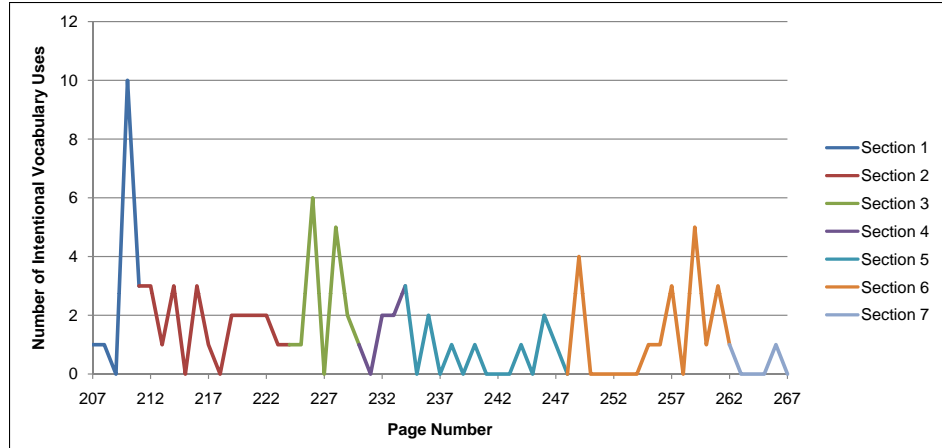


Figure 22: Occurrences of Intentional Vocabulary in “The Copycat Project: A Model of Mental Fluidity and Analogy-making” [98]

words, the majority of the comparison between *Copycat* and other systems is carried in the technical language. For instance, notice the difference between the following quote and the example shown in the beginning of this section:

Copycat’s drives toward abstraction and systematicity were just described above. But note that Gentner’s definition of “abstraction” [order of a relation] is not the same as the notion in Copycat of “conceptual depth.” In Copycat, there is no logic-based definition for conceptual depth; rather, these values are assigned by hand, with quite high values sometimes going to concepts that Gentner might call “attributes” (such as alphabetic first, which could be seen as an attribute of the letter “a”).

The reason for this contrast reveals the “awkward” role of intentional narrations play in the practice of AI. Despite the importance of intentional narrations, their subjectiveness and richness portray them as the antithesis of the scientific practice. As a result, they are often neglected, denied of significance, even by their creators. Hofstadter and Mitchell, on the one hand, painstakingly constructed multiple layers of narrations to portray the human-like behavior of their system — their codelets “scout” the unknown territory and “cry out” for attention. On the other hand, when

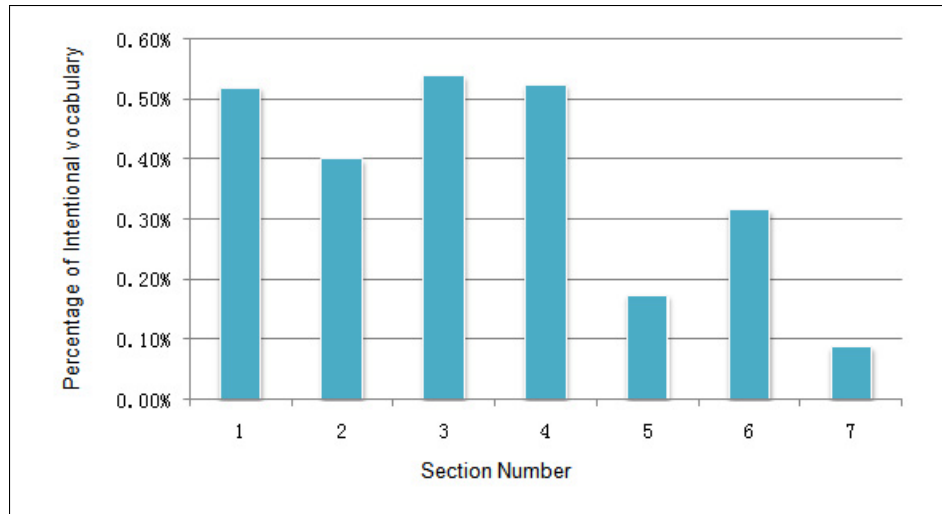


Figure 23: Density of Intentional Vocabulary in “The Copycat Project: A Model of Mental Fluidity and Analogy-making” [98]

confronting another system, they shy away from their discursive machine and turn completely to the “hard-and-fast” facts of the system’s technical operation, hoping that they will not be accused of the “epidemic” of “wishful mnemonics” (Section 3.2.2). It is remarkable when the authors wrote:

Much the same could be said for Copycat: an outside observer is free, after a run is over, to “parse” the run in terms of specific, discrete processes, and to attempt to impose such a vocabulary on the system’s behavior; however, that parsing and labeling is not intrinsic to the system, and such interpretations are in no way unique or absolute, any more than in a basketball game. In other words, a long sequence of codelet actions can add up to what could be perceived a posteriori and by an outsider, as a single coherent drive towards a particular goal, but that is the outsider’s subjective interpretation.” [97, pp.50]

Here, Hofstadter and Mitchell rightly acknowledge the multiplicity of system intentionality. However, they conflate the authors’ active narration of the system with

users' interpretation. Their own example illustrates the pervasive and yet illegitimate existence of intentional narrations in the practice of AI.

4.5 *Ideological Analysis*

Given his enthusiasm for, and knowledge of, art, music, and mathematical beauty, as illustrated in Hofstadter's first book *Gödel, Escher, Bach*[93], it may be a little surprising to many people that he chose alphabetical strings as the domain of his program. Admittedly, AI has a history of using simplified microworlds to solve otherwise overly complex problems, and the mathematics and physics trained researcher is certainly familiar with this reductionist approach.² Hofstadter's position in AI is well captured in his interview with the magazine *WIRED* in 1995 [110], shortly after his second book was published.

...you catch me in a tricky dilemma here. I want any computer program that my students and I work on to delight me with its cleverness. I want it to outclever its programmers. But at the same time, if after 10 or 20 years of work my program composed a great novel, made a series of great mathematical discoveries, or wrote lots of great one-liners, I would be terribly distressed. I feel the human spirit is infinitely more complex than anything that we're going to be able to create in the short run. And if we somehow did create it in the short run, it would mean that we aren't so complex after all, and that we've all been tricking ourselves.

Notice that Hofstadter's worldview, although close to the weak AI hypothesis, is still different from it. As discussed earlier, the weak AI hypothesis holds that computers may appear to be intelligent in certain circumstances, but they do not

²A microworld in AI is a limited domain of problems that appear to require intelligence to solve. The most famous microworld is the "blocks world," used by many AI researchers including Winograd. Compared with the complexity of real world, a blocks world consists of blocks on the table, waiting to be stacked in certain ways.

have a mind, intentionality, or consciousness. The *Copycat* project, however, is more ambitious. Its authors aim towards teaching the/a computer the “core of thinking” [110] — analogy — in a psychological plausible way. On the other hand, as Hofstadter has expressed in various contexts, he strongly believes that certain human qualities, “something deep” at our “core,” that manifest itself through music and writing, are not achievable by computers.

This chapter has illustrated a close reading of a full-fledged analogy making system, *Copycat* in order to demonstrate the utility of the AI hermeneutic network. Different from the existing methods of analyzing computational systems such as AI/computer science analysis and human computer interaction evaluation, an interpretive approach was adopted here. The primary corpus used in the analysis was the technical literature published by the authors of the system. In addition, secondary material such as the authors’ non-technical articles and interviews in the mass media was also used for references. Based on a method informed by classical hermeneutics, a close reading of the system was carried out on three levels: the technical, social and cultural context during which the system was developed, content analysis, and ideological analysis.

First, the context analysis has revealed that the *Copycat* project was historically situated between the two “AI winters,” a transitional period when the GOFAI paradigm was slowly giving way to various approaches of AI. This had an impact on both the algorithmic deployment and discursive strategies *Copycat*’s authors took. The main author’s (Hofstadter) connection with other actants (e.g., Marvin Minsky, funding agencies) was also discussed.

The primary focus of the study, among the three levels, was the content analysis. Based on an empirical analysis of the technical literature (the main corpus), I have

identified the intentional vocabulary’s utility for connecting the two languages simultaneously at work in the authors’ narration of their system — one technical and the other intentional. Through two specific examples of the *happiness* level of codelets as well as the *scout* codelets, this chapter demonstrated how the authors of *Copycat* use intentional vocabulary to support the technical requirements and the intentional narration, both needed to justify the system’s operation and significance. Next, a pattern of when the authors of *Copycat* switch between the technical and intentional language for different purposes was presented.

In the last step of the ideological analysis, I contextualized the technical literature under study in the context of the main author (Hofstadter)’s philosophical view on AI. The discovery of the two languages and the role of intentional vocabulary to connect them based on empirical data matches with the theoretical discussion of the discursive machine previously presented in Chapter 3. It further proved the author’s active role in narrating system intentionality. Most of the above observations have escaped the analysis of traditional AI (computer science) and HCI frameworks.

So far, we have answered the first research question of the dissertation: Where does system intentionality come from? The previous two chapters (Chapter 3 and 4) have argued, both theoretically and empirically, that authors’ narration and users’ interpretation are the pivotal elements in the formation of system intentionality. The discursive and interpretive nature of system intentionality hence allows us to think of new ways of designing and creating intentional systems. The next two chapters hence focus on the second research question: In light of the theoretical findings, how do we design novel intentional systems?

CHAPTER V

AGENCY PLAY AS A SCALE OF INTENTIONALITY

You are playing a young brave warrior in a sparse landscape. In order to save a loved one, you must slay towering mythical behemoths, one after another. After slaying a creature by scaling it and plunging a sword into a tender point, the giant drops to its knees in melancholy slow motion. You run, jump this way and that, trying to avoid black veins of dark force rising from its body to seek you out. Inevitably, they pierce you and the world turns black ...

— A scene from video game *Shadow of the Colossus* [198]

The first half of the dissertation has focused on authors' narration of system intentionality. The rest will propose new design possibilities and strategies to celebrate system intentionality, not as the AI dream, but as playful and expressive new experiences for the users. What factors should we take into consideration while designing intentional systems in expressive, entertainment, or even educational domains? Is there any design space that has not been explored sufficiently by existing approaches? What would an expressive intentional system be like? These are the main questions that the next two chapters intend to answer. This chapter offers two new design constructs, *scale of intentionality*, and *agency play*, both of which were developed jointly with Fox Harrell [220, 84]. Scale of intentionality is a novel approach to turn system intentionality into a meaning-making mechanism. Special attention is paid to one instance of the scale of intentionality, *agency play*, which explores system agency as a potential narrative device that can be used to convey meanings in addition to

the story content. The term system agency and its close connection to system intentionality are introduced in Section 5.1, where the limitation of narrowly defined approach to agency in human computer interaction and interactive narrative works is also discussed. Next, Section 5.2 traces cross-disciplinary notions of agency from various fields, and proposes an understanding of user agency as situated in the context and goals of the narrative work at hand. Agency over some dimensions of control such as moving a player character can be pitted against agency in determining plot outcomes. The following Section 5.3 proposes the new model of agency play in which control over a broad set of possible dimensions of user and system interaction becomes a mechanism to convey meaning and express ideas. Toward this end, we will look at cases in interactive narrative media and in games with strong narratively structured content or rich story worlds. Finally, Section 5.3 explores various strategies of agency play based on a new taxonomy of the different layers of agency. Although most of the discussion here will focus on the domain of interactive narrative and its derivatives in computer games, the core of design concepts raised is applicable to a larger realm of intentional systems and digital media.

5.1 Scale of Intentionality and System Agency

In this section, we will first go over part of the history from the early development of film, for it contains valuable lessons for intentional system designs. Based on this history, a novel design concept of scale of intentionality is proposed to turn system intentionality into a new expressive resource of intentional systems. We will pay special attention to one particular way that system intentionality can be manifested in interactive narrative works, that is, system agency.

5.1.1 Scale of Intentionality

When the Lumière brothers first experimented with motion pictures during the mid 1890s, the films that they produced were without exception shot from a camera



Figure 24: Lumière Brothers' Early Film "Workers Leaving a Factory," screened publicly in Paris in 1895

fastened to a stationary tripod. Events unfolded by people and props moving in front of the camera, whereas the camera stayed in the same position (Figure 24). This setting was unquestionable to many people at that time because it paralleled a more familiar setting — theatre. In traditional theatric performances, a spectator is expected to sit in a fixed position while she watches actions taking place in front of her.

It was not until almost two decades later when this setting was fundamentally challenged by D. W. Griffith. Film scholar Tom Gunning [75] provides a list of new elements that Griffith introduced, including the full shot, the medium shot, the close-up, the pan shot, the moving camera, the spot-iris, the mask, and the fade. Despite Griffith's extremely brutal discrimination against American Americans, these new inventions not only freed camera from being a passive instrument needed for the filmmaking process, but also contributed to the emergence of film as an expressive language. D. W. Griffith was referred to by Gunning, among other, as the master of

the syntax of film narration, who developed film from “cinema of attractions” to a narrative form.

This era in the history of film is informative for designing new intentional systems, especially in the domain of interactive narratives, for two main reasons. First, it extends the literary distinction between narrative content and discourse to the new form of narrative. The “same” story content, assuming such platonic concept exists on its own, can be turned into different films, depending on the camera work and other “discourse” elements used by different directors. Similarly in interactive narrative, the various representation and amount of system intentionality can potentially be used to carry meanings independently from the story content.

Second, in addition to pointing out the discourse level, Griffith developed and demonstrated different ways that the *variation* of elements at this level can be used to exaggerate, propagandize, or even subvert the story content. This is an area that has not been systematically explored by existing intentional systems yet. Most of intentional systems discussed in this dissertation maintained a static amount of system intentionality — certain systems (e.g., *Voyager* and *Aaron*) are more intentional than others (e.g., *Photoshop*), but each system’s intentionality mostly stay the same over time. What if we can change this status quo?

The broad design concept *scale of intentionality* proposes to dynamically adjust system intentionality along a scale through various discursive strategies. It explores novel ways in which the variation in system intentionality is used to convey meanings and express ideas. In the domain of interactive narrative, system intentionality manifests itself through several aspects, especially through characters’ agency and disposition. For instance, our *Memory, Reverie Machine (MRM)* system presented in the following chapter invokes a scale between narrating highly user-controlled character behaviors directed by the user’s actions and desires within the story world and highly autonomous ones that exhibit situated “aboutness” regarding the system’s

agency within its domain of operation. The ways in which the system’s agency can be manipulated to make meanings will be discussed in full length in the rest of the chapter. The *MRM* system also dynamically adjusts the degree to which the character’s dispositions and emotions are generated and are visible to the user. On the higher end of the scale of intentionality, the character expresses his mental states and emotions through memories and daydreams, whereas on the lower end such mental activities are completely omitted from the story generated by the system. The use of disposition in the scale of intentionality will be illustrated through our *MRM* system in the following chapter.

5.1.2 System Agency

The design of intentional systems often involves the choice of temporarily suspending users’ control over the system. For instance, the human jazz improvisers do not have direct and complete control over the *Voyager* system in the way that a user of *Photoshop* does. The design choice of dialogic exchange between human and computer that helps to achieve *Voyager*’s status as an independent player is, however, is at odds with the traditional value of human computer interaction (HCI) design paradigm. As a direct descendent of the “computer as a tool” perspective (Section 2.1.2), a major part of HCI is devoted to designing computer systems are are transparent and easy to control [161, 40].

For example, Ben Shneiderman, one of the most influential researchers in HCI, posited the influential concept “direct manipulation” for interface design. It focuses on interface representations that can be transparently manipulated and users’ abilities to apply intellect directly to the task at hand [194]. In his well-known debates with Pattie Maes on direct manipulation vs. deploying software agents as interface strategies, Shneiderman stands strongly behind the former because it “gives users the feeling of being in control and therefore they can be responsible for the decisions

they make” [195]. He further argues that “I make the basic assertion that people are not machines and machines are not people. I do not think that human-to-human interaction is a good model for the design of user interfaces.”

Shneiderman’s depiction of the feeling of “being in control” can be partially described by the philosophical concept of agency. The philosophical concept of agency generally refers to the capacity of an agent to act in a world. This construct is closely related to intentionality, for the agent’s capacity is inevitably influenced by her mental states of beliefs, desires, etc. It is true that theorists have placed different levels of emphasis on intentionality in their definitions of agency. For instance, Giddens [70, pp.56] acknowledges “the intentional or purposive character of human behavior,” while emphasizing “[s]uch intentionality is a routine feature of human conduct, and does not imply that actors have definite goals consciously held in mind during the course of their activities.” Others believe that it is “the strong role of active (though not necessarily fully ‘conscious’) intentionality in agency that ... differentiates agency from routine practices” [163, pp.136]. The details of these different approaches are not essential to this chapter and has been elaborated in Ortner’s work [163, pp.134-54]. The most important message here is that agency is one of the ways that intentionality can be externalized through visible actions that are cognitively and emotionally directed *towards* certain things and/or purposes. As Ortner suggests, “intentionality in agency might include highly conscious plots and plans and schemes; somewhat more nebulous aims, goals and ideals; and finally desires, wants and needs that may range from being deeply buried to quite consciously felt.” This close connection between agency and intentionality is of particular relevance to apply scale of intentionality to interactive narratives, in which agency has already been widely discussed [154].

If HCI is seen as focusing on users’ agency, the design concepts described in this chapter introduce the construct of system agency and propose a new balance between users’ and systems’ agency. Subsequently, a system’s agency to act independently is

not to be seen as a flaw because it perceivably takes the control away from the user. Instead, carefully designed suspension of user agency under certain circumstances can be a meaning making mechanism, as shown in the examples of interactive narratives later in the chapter, or a political strategy to call into questions issues of “control” and “diversity” as in George Lewis’ case. In other words, system agency is complementary to user agency.

Overall, *System agency* is one of the ways that system intentionality can be externalized. System agency can be manifested as the capacity of a computational system to act independently and autonomously, performing actions or generating events while being designed and described in ways that facilitate intentional readings of these actions and events. When the level of system agency is high, a system’s process is typically opaque with higher algorithmic complexity and it may display more complex, human-like behaviors (Section 1.2).

The notion of system agency leverages the discourse accompanying classic AI goals of building autonomous intelligent agents whose behaviors reveal human-like directedness toward the world. The term “system agency” does not imply, however, a dream of the strong-AI-styled system autonomy. Similarly to system intentionality, system agency is a result of human interpretation in light of a set of situated social circumstances, not as a property intrinsic to a computational system itself. The term “system agency” provides a shorthand to describe human interpretation of properties of the system behavior and capacity, specified by the story author and authoring system designer. More importantly, it suggests that system agency needs to be considered in parallel to human agency. The name is selected to call attention to the inseparable nature of the human and system agency, as well as to sketch the possibility for skillful manipulation of the relationship and dynamics between them toward expressive ends.

5.1.3 System Agency in Interactive Narratives

The description in the beginning of the chapter depicts a narrative scene from the console video game *Shadow of the Colossus* [198]. Some readers may wonder why the game design allows the player to perform actions immediately after slaying each monster when the resultant fall into blackness is unavoidable; after all, many games would use a pre-rendered, non-interactive cut-scene instead. Since the ending of this scene is pre-determined and beyond user's power to change, is this "false" illusion of control a mistake overlooked by the designers of the game?

Digital media afford more active roles for users than traditional narrative forms in various ways. Murray [154] identifies this phenomenon of increased user agency and claims it as one foundational property of digital media. She also categorizes various ways in which user agency is externalized in interactive narrative works. Examples of user agency include spatial navigation, problem solving, incorporating gameplay actions within narratives, and traversing links in hypertext narrative.

Murray's theory has influenced many practitioners who set out to explore new narrative possibilities brought by digital media in which agency plays an important part. However, many attempts have been based on an overly simplified understanding of agency, that is, agency as free will of players. This interpretation inevitably leads to an obsession with the idea that "the more agency, the better." This phenomenon is exemplified in notions such as "full reactive eyes entertainment" or F.R.E.E., a concept proposed by game designer Yu Suzuki during the development and marketing of the game *Shenmue* [191], which aspired to allow a character to freely interact with every aspect of her environment. Yet, despite the high degree of free will to interact with objects in such a game, the slow pacing of its plot was consistently viewed as a detraction from the game, perhaps with the high degree of freedom detracting from meaningful narrative development, resulting in player boredom.

This chapter is especially geared toward computational narrative systems that actively generate stories, story worlds, and/or dynamically alter elements of narration, using AI algorithmic and knowledge structuring approaches. Such systems embody a wide range of types of system agency that often fall into two categories, the second a subcategory of the first: 1) system agency as narrative generation and constraint and 2) system agency as character control. The first type refers to the capacity of system as the “narrator” and/or “author” to affect what the story is (narrative) and how the story is told (narration) through control of events and objects, and their presentation in the story world. Although computational narrative generation systems have traditionally focused on narrative presentation and structuring of events such as in [142], recent work such as [150] has emphasized more complex variation in the narration of stories. The second type is system’s capacity to control events through control of *characters* in the story, often computationally encoding both internal states and external behaviors of characters. This type of system agency often relates to believable agents design, and gives rise to the illusion of intentional agents acting in the story. As shown in the next section, when both user and system compete or collaborate in control of the same characters, a new range of opportunities are enabled to study and implement new ways of designing agency toward expressive and aesthetic ends.

Context is indispensable to understanding of system agency in interactive narrative works as it simultaneously constrains and enables the nature and interpretation of user actions. Computational systems play a central role in establishing meaningful contexts in interactive narrative works, ranging from presenting hypermedia experiences of many electronic literature works, such as the lauded text-based animation “*Dakota*” [219], to selecting story beats in Mateas and Stern’s interactive drama *Façade*.¹ The capacity of the computational system to modify the story world and

¹*Dakota* is an interesting example of system agency, even though the Flash-based system does not fall into the prototype of intentional systems. Not only does the system completely deprive the user of her agency to interact with the piece, a more surface reading of system agency, but also it

provide affordances for users actions is system agency.

Although Shneiderman’s concept of direct manipulation and other related HCI design principles are certainly valuable for user interface design, it certainly has its limitations when applied to intentional systems particularly in entertainment and aesthetic domains such as interactive narrative. A user’s actions and their consequences in such contexts need to carry purpose and significance. Computers (and other artifacts) can contribute to the perceived meaning of human actions to take place, once the computers are not seen merely as instruments.

5.2 A Situated Approach to Agency

The central argument of this section is that solely free will oriented, more is better, models of agency are outmoded and should be expanded with an understanding of agency as an expressive resource that can be varied along many dimensions to result in meaningful and aesthetic effects in interactive narrative works. Formerly, such as in the heyday of hypertext fiction or adventure gaming, user agency was novel in the ability to traverse multiple paths through, or even become a character within, an interactive narrative work. Full player agency became a type of holy grail in the construction of interactive narrative works. However, the field has matured beyond that stage, and the deployment of particular affordances for user agency is now a stylistic choice carefully manipulated over the course of a user’s experience to facilitate the goals of narrative system designers. The text of the rest of this chapter is mostly a reprint of Harrell & Zhu’s paper “Agency Play: Dimensions of Agency for Interactive Narrative Design” [84], presented at the AAAI 2009 Spring Symposium on Interactive Narrative Technologies II.

Long before being adopted in the digital media theory, apprehending the concept of agency has been a central challenge in various disciplines including philosophy

subverts the established conventions and expectations of reading.

and anthropology, as well as political activism and critical cultural theory. The focus here is addressing a set of theories that allows us to scrutinize user agency in relation to the digital technologies (computers, algorithms, etc) that go hand-in-hand with it. The work of Andrew Pickering in the sociology of scientific knowledge influences us in its distinction between human and material agency, articulation of a “dance” between them, and focus upon a type of user agency as knowledge of how to perform manipulations in a particular conceptual system (e.g. how to perform elementary algebra manipulations) [173]. The actor-network theory of Michel Callon and Bruno Latour has been influential in our practice in the way it incorporates both human and non-human actors [124]. Furthermore, in her review of various accounts of agency, Laura M. Ahearn identifies three major trends of which the concept has been conceptualized over the past few decades, namely “agency as free will,” “equating agency with resistance” and “the absence of agency” [4]. Although the prevailing view in the domain of interactive narrative aligns itself with the first trend, the recent years have witnessed increasing number of experiments drawn on the other two accounts. The end of this section proposes an approach to agency that is situated in particular social and narrative contexts and lends itself to a range of expressive applications in interactive narrative works.

5.2.1 A Dance of Agency

Andrew Pickering focuses on a duality that he describes as existing between “human agency” and “material agency” [173]. Human agency centers upon the intentionality (aboutness) and actions taken by humans upon the world. In particular, as he is interested in the types of agency exhibited by scientists in the practice of knowledge production. He describes human disciplinary agency, recognizing and knowing how to use/perform “a series of manipulations within an established conceptual system.” In contrast, Pickering also describes “material agency” as the idea that the world is

“continually *doing things*, things that bear upon us not as disembodied intellects but as forces upon material beings.” It is a view that sees science as an array of forces that humans typically apprehend through the use of machines. Pickering defines a “dance of agency” as the process where humans attempt to apprehend the agency of the material world through the mediation of artifacts, and that material world both yields to, and resists, human apprehension. Despite his focus on scientific practice, the notion that there are symmetrical notions of human and material agency engaged in a dialectical dance parallels our constructs of human agency and system agency in AI and interactive narrative practices.

5.2.2 Agency as Free Will

The prevailing approach to user agency in interactive narrative systems treats the term as synonymous with users’ free will. A straightforward example of such a treatment occurs in rhetoric of free spatial navigation, where a user is allowed to explore digital environments at will, without being bound to visit locations in an order predefined by author. Such environments include map locations in video games or text-based interactive fiction (IF) labyrinths. In describing such environments, critics often describe possibilities of spatial exploration as cases of user agency. In articulating different categories of interactive narrative, literary theorist Marie-Laure Ryan captures this conception of agency in her description of works providing what she terms “internal-exploratory interactivity” [181]. She describes these as systems where “the user exercises her agency by moving around the fictional world, picking up objects and looking at them, viewing the action from different points of view, investigating a case, and trying to reconstitute events that have taken place a long time ago.”

The above examples refer to a user’s capacity to take actions and observe their results. As Ahearn reminds us, this capacity is deeply connected to what philosophers

call *action theory* in that each of the above examples requires certain concomitant mental states, such as “intention” [33] and “presence of the self” [192]. As action theorist Segal puts it “[h]itting a ball is an action, falling down a flight of stairs is not. A theory of action seeks, among other things, to explain the distinctions we make.” Increasingly powerful hardware and more robustly interactive software have allowed designers to explore the decisions users can make at will in forms unimaginable to early-day developers. For example, in the Xbox game *Star Wars: Knights of the Old Republic*, the way a player controls her character through the story determines its moral disposition (whether it aligns with the light or dark side of the force), the set of new skills it can learn, as well as the branches of storyline which will be taken [132].

However, an unchecked focus on free will may lead to unanticipated, and largely undesirable (though perhaps also transgressive) consequences. In massively multiplayer online role-playing games (MMORPG), grief players perform actions not to advance game goals or for narratively oriented fulfillment, but to intentionally aggravate and harass other players. They kill other players, steal weapons and coins, and even form virtual mafia, all in the pursuit of their own enjoyment and free will version of “agency.”

The example of griefers echoes Ahearn’s warning to us. “The main weakness in treating agency as a synonym for free will,” she argues, “is that such an approach ignores or only gives lip service to the social nature of agency and the pervasive influence of culture on human intentions, beliefs, and actions.” When adopted in the domain of interactive narrative, this pitfall often transforms into the over amplification of users’ freedom to act however they want, while overlooking the importance of meaningful constraints and conditions in the context of the story.

5.2.3 Agency as Resistance

“Agency as resistance” characterizes many works in feminist theory and subaltern studies, in which traditions of social resistance of the past and present are called into attention. Many feminist theorists agree that, in order to demonstrate agency, a person must resist the hegemonic patriarchal status quo [1]. This form of oppositional agency has gradually been adopted by some users/artists/hackers of digital worlds. In 1999 Sonya Roberts released her *Female Skin Pack Excerpts*, a series of female texture map for the original *Quake* avatars, because the game designers neglected to provide a female protagonist. The eerie composition of a female skin on a muscular male figure embodied a form of resistance to power.

Oppositional agency also finds its way through the voices of “protesters” in virtual environment. *Second Life* users have successfully pressed their developer Linden Labs to alter the regulations in the Internet-based 3D virtual world in various cases, most prominently two events in 2003: a virtual tax revolt and agitation to allow people to retain IP rights. It is useful to include such notions of agency because they relate user action to broader social, political, and cultural contexts both within and outside of the story worlds of particular interactive narrative (and related) systems.

5.2.4 Absence of Agency

Another approach to agency is well articulated by Michel Foucault’s work on power [60, 61]. An extreme reading of Foucault is that omnipresent impersonal discourses so thoroughly pervade society that no room is left for anything that might be regarded as agency, oppositional or otherwise [4]. After playing some games produced by Ian Bogost’s company Persuasive Games, such as *Airport Insecurity* and *Disaffected!*, it is difficult not to question the existence of agency, both inside the games and out, when a user is pushed to accomplish tasks either specified by game mechanics or more pervasive commercial bureaucracy and protocol in real life.

5.3 *Agency Play as an Expressive Tool*

In light of the above, our notion of agency needs to reconcile valuable insights provided by the various perspectives on agency. Any unilateral definition of agency is inadequate. An actor’s agency is mediated through situated rules and resources. Structuration theory addresses such negotiation between social structures and human actions. Her actions are simultaneously constrained and enabled by the very social structures that those actions serve to reinforce or reconfigure [70, 4]. As Ahearn proposes, agency refers to the *socio-culturally mediated* capacity to act. In digital environments, a user’s power to take meaningful actions is mediated through the structure provided by the computational system as well as the socially situated interpretation of actions rendered by the user. A system’s capacity to afford certain actions, impose certain constraints, and reward certain behaviors clearly has great effect on user’s agency. Even though games such as *Shenmue* are often accompanied by strong rhetoric that a player can do anything they like and their actions will impact the gameplay, the system structure imposes a strong value system that shapes the player’s agency, which, in turn, may reinforce or allow for critique of that structure. User agency is thus situated materially in the system affordances and interpretively in the context of use.

User agency also provides an aesthetic experience and needs to be appropriate to its narrative context. A user’s capacity to act and make distinction does not necessarily entail narrative consistency. Fortunately, the area of interactive narrative is largely comprised of works that incorporate what Murray [154] calls the balance between “dramatic structure” and “player freedom” [138, 78]. Aspects of agency that have been commonly underexplored, however, include the ways that dynamically changing the scope, nature, and degree of user agency during execution can serve the expressive goals of an interactive narrative work. As a form of cultural production, interactive narratives are created by human authors to convey meaning. What is

significant sometimes is not *what* the story is, but rather *how* the story is told. The use of agency is one of the channels for digital authors to express themselves.

An equally expressive use is to limit or even temporarily eliminate user agency to convey a certain message, such as the sense of confinement or helplessness. Such strategy nevertheless could be risky because historically the lack of agency, as freewill in particular, has been associated with the computational simplicity of the system or deficiency of the design. The use of cut-scenes in video games is a good example. For decades the game industry has incorporated cut-scenes where user freedom is temporarily suspended. Although scholars have pointed out the narrative utility of non-interactive cut-scenes, e.g., to advance plot and introduce characters [108], recent trends encourage designers to incorporate player interaction in these cut-scenes regardless of its content, for example in the introduction scene of *Half Life 2*. It is sometimes easy to forget that the power of narrative systems is to provide units of meaningful work, not mechanical interaction [14]. As argued in the rest of the chapter, computational techniques can be used to tune user agency according to the story context.

5.3.1 Agency Play

Agency play engages multiple aspects of agency during runtime in expressive ways that reinforce the authorially intended meanings of the narrative. Agency play is not merely strategically limiting user control. It focuses on leveraging the relationship between the user and system in order to create a story world that is meaningful and engaging for users to participate in. Incorporation of a range of agency effects in interactive narrative is similar to the step that film makers took last century when they discovered that camera angle, framing, and take length were all effective storytelling mechanisms, not just coincidental properties of the medium. This section defines promising dimensions of expressive use of agency and provides illustrative examples.

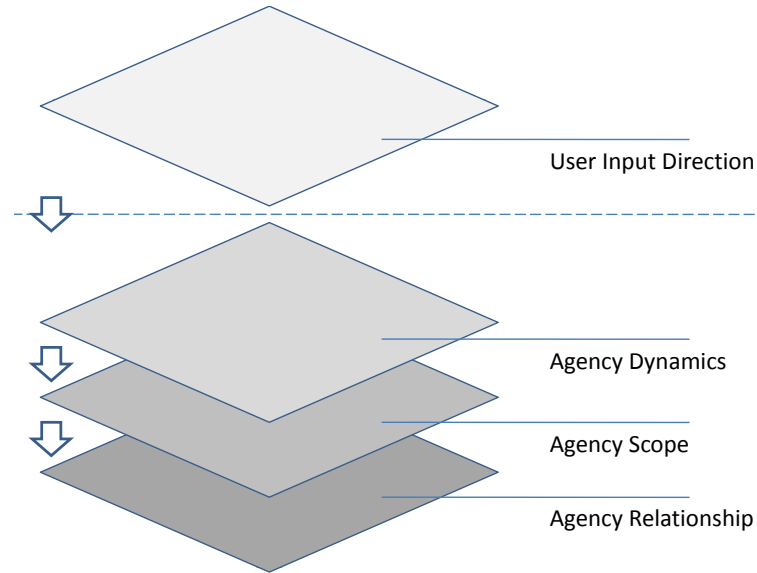


Figure 25: Dimensions of Agency (Arrows Indicate the Direction of Influence)

Accordingly, this section offers the following layers of agency play, each of which can be used expressively to convey meaning:

- *Agency Relationship*: User actions and system actions operate in relation to one another and can vary in relative magnitude and degree of dependency on one another (e.g. an inverse relationship or independent operation).
- *Agency Scope*: Results of either user or system actions may have immediate and local impact (e.g. turning a character left or right) or longer term and less immediately apparent results (e.g. a series of actions may determine narrative structure itself).
- *Agency Dynamics*: The relationship between possible user and system actions, and their scopes, can vary dynamically during runtime.
- *User Input Direction*: The user may establish a pattern of input that directs agency dynamics and/or agency scope.

Figure 25 illustrates how the dimensions of agency relate to each other and how each dimension mediates the ones below it. The dimension of User Input Direction is

separated graphically because it directs the deployment of effects of the levels below it, but is not itself constituted by the layers below it. These layers below are meant to provide a vocabulary for more precisely describing such varying types of agency, and how they can be manipulated toward expressive effects.

5.3.2 Agency Relationship

Agency relationship is the fundamental dimension of agency play, and refers to a function, always situated in a particular context, describing the relative magnitude of and dependence relationship existing between:

- a set of actions allowed (by the system) to be executed by a user
- a set of actions allowed to be executed by the system.
- user desire or need to perform actions,
- user sense of meaningful possible actions, and
- a range of possible user interpretations of actions.

Focus is given to the first two aspects of the function above, with the understanding that the latter three aspects determine the expressive qualities of the agency relationship. *Relative magnitude* refers to the relative degree of possible user actions to system imposed constraints, especially constraints upon actions that would usually be enacted by the user in the story world.

Dependence between player and system control over actions operate vary greatly. Primary models include cases where user and system agency are 1) independent, 2) inversely dependent, 3) interdependent. Of course, standard characters that are transparently controllable by a human player and characters that are controlled solely by the system operate independently. Independent user agency in interactive narrative works usually refers to the ability of the player to take meaningful actions in a story world. The ability to cause a character to move, acquire artifacts, and interact

with other players or non-player characters (NPC) reflects this sense of player agency. Yet, user agency can also operate along dimensions outside of the story world such as in-game camera control. The “meaningfulness” of player actions most often arises from the degree to which a player’s actions express intentionality, or aboutness, in the story world, and the ability of player determined actions to affect the world and progress in a narrative. Thus, control over character actions that do not have any significant narrative effect are commonly described as providing the player with a low degree of agency. Simultaneously, moving a camera in a proper way may reveal to the player a necessary object in the game world for progressing the narrative, an example in which a non-story world action provides a meaningful story world result. As argued above, the notion of *meaningful* agency relies upon situated construal of possible user actions.

Independent system agency can refer to narration of the system as being capable of autonomously carrying out human-like actions. For example, Rafael Perez’s system MEXICA has been described as an automatic storytelling system[168]. MEXICA itself exhibits system agency by doing the “telling.”²

The IMPROV system by Ken Perlin and Athomas Goldberg, which was developed to allow for the scripting of semi-autonomous characters that expressed their own behavior-driven goals and states, provides a good case with which to describe inverse dependence [171]. For example, one could say that the more an Improv-based character can take action in a story world without direction of a player, the more system agency and less player agency is exhibited. The 2005 strategy game *Civilization IV*, on the other hand, represents a different correlation. When the agency of the player increases in the game as she or he gains more resources (money, weapon, technology, etc), so does agency of the system because it controls more non-player

²Pérez y Pérez himself does not always deploy such intentional narratives in articulating the system’s operation, describing it as a model for producing story frameworks in his website www.rafaelperezyperez.com/MEXICA.

characters (NPCs) with increased capability, and the system must coordinate a more complicated set of game world events at large. This represents a type of interdependence. Likewise, a system could implement a subset of playable characters that are semi-autonomous, requiring only high level direction from a player. In this case the control of character action also displays an interdependent relationship. All three types of dependency articulated above are meant only to sketch useful points along a range of possibilities, rather than to exhaustively list every possible type of dependence between user and system control of situated action.

5.3.3 Agency Scope

The concept of agency scope describes the impact and narrative focus of user and system actions, ranging from immediate and local impact, such as spatial navigation ability, to less immediately apparent but more global results, such as shaping the narrative structure itself. Either side of the agency scope spectrum can be used effectively to convey meanings in addition to the actual narrative.

We have already seen how *Shadow of the Colossus* deploys a high degree of local player agency with a low degree of global player agency to render the sense of fate and helplessness. An opposite practice can be found in the interactive narrative documentary *Terminal Time* [136]. In this work, users only have very low local agency through the one-dimensional control mechanism (volume of the clapping sound) with very few multiple choice questions. However, the variability in the generated output in terms of both media elements invoked and rhetorical model is great. The contrast, in this case, between low local player agency and high global player agency can be read as a commentary on pervasive power of ideologies.

5.3.4 Agency Dynamics

The nature of a given agency relationship and the scope of agency impact can vary *over time*. In this case, we say that there is a play of agency dynamics. If these

dynamics are orchestrated in order to express a theme such as the increasing emotional maturity of a character, then agency play has become an expressive resource varying according to runtime aesthetic dictates. Stories that contain fixed level of agency relationship and agency scope throughout runtime, which could be a conscious and expressive design decision in its own right, have static agency dynamics.

One space to explore the dynamics between player agency and system agency is through *semi-autonomous player characters* (SPCs). In the domains of interactive narrative and gaming, characters are often categorized as player characters (PCs) or NPCs. PCs are often presented as avatars entirely controlled by players, whereas NPCs embrace system autonomy and are not usually subject to player command. Although most characters fall into one or the other category, some incorporate traits of both. For instance, some PCs may convey their impatience by foot-tapping or may smile when receiving power-ups. Our current work [221, 220] explores how to algorithmically adjust agency dynamics in SPCs with a goal to express our view of the relationship between human and machines(AI). In the *Memory, Reverie Machine* project (formerly called the *Daydreaming Machine*) that is described in the next section, both the relationship between player and system agency and their respective scope level changes as required by the story and character development over time.

5.3.5 User Input Direction

All the above levels of agency can be directed by user input. For example, in *Pac-Man 2: The New Adventures* the player does not directly control the PC, but can direct his attention toward certain objects or tasks (which the character sometimes refuses). This is used expressively to provide a sense of personality to the player character, at the same time as providing a necessary constraint on possible actions. In *Terminal Time*, described above, the player interaction to determine the slant (e.g. feminist or techno-centric) is minimal, the audience collectively answers a set of multiple choice

questions. However, the user input is what offers players the strong sense of global agency in the piece.

The model presented above is a preliminary effort to carefully present a model of agency that includes often overlooked agency phenomena, towards the development of theory for design of interactive narrative systems. The account of all four dimensions above certainly is not comprehensive, and the examples do not cover the entire area of expressive possibilities. Each is an area ripe for further exploration. Our modest goal here is to present a new approach to considering the role of agency in interactive narrative works and to provide vocabulary for the discussion.

In this chapter, we have introduced the general design concept of “scale of intentionality” that explores the use of system intentionality to convey meaning. Special attention was paid to one particular instantiation of this design concept, agency play, in the domain of interactive narratives. First, critical attention has been called to certain narrow understanding of agency in interactive narratives. After raising the construct of system agency, this chapter has presented several socially situated definitions of agency in contrast with a popular perception of concept as simply free will. In opposition to a subsequent common assumption of “the more agency the better,” this chapter argues that narrative agency is contextually situated, distributed between the player and system, and mediated through user interpretation of system behavior and system affordances for user actions. Next, a closer examination of agency allows us to break it down into four dimensions, whose narrative effects are illustrated through examples. In addition to agency play, the scale of intentionality can also be manifested through adjusting the ways in which a character’s disposition is narrated. Various deployments of scale of intentionality will be further explained in the next chapter in the context of our interactive narrative system *Memory, Reverie Machine*.

CHAPTER VI

MEMORY, REVERIE MACHINE: A CASE STUDY

I wonder if a memory is something you have or something you’ve lost.

— Woody Allen, *Another Woman* [5], 1988

Intentionality and agency are crucial links that connect an individual’s mental states to the material world and society around. Actions are certainly a major way to manifest these constructs, as shown in prior discussions regarding both real life scenarios and digital simulations in interactive narrative works. Insufficient attention in the interactive narrative community, however, has been paid to one’s mental and imaginative activities. Memories, reveries, and daydreams not only are another crucial locus where one’s beliefs, desires negotiate with the world which they are directed at, but also offer a rich, expressive domain for artists, writers, and designers to explore.

This chapter resumes the discussion of *The Memory, Reverie Machine* (formerly named as *Daydreaming Machine*)(*MRM*), a text-based interactive narrative work co-developed with Fox Harrell, from the introduction chapter (Section 1.4). *MRM* was built on Harrell’s theoretical framework for imaginative multimedia discourse [81, 82] and technical framework for interactive and generative discourse instantiated in the *GRIOT* system (Section 2.4.2). The author of this dissertation has contributed the narrative content and several technical extensions to *GRIOT*, necessary for *MRM*.

Our major goal for the project is two-fold, one reinforcing another. First, as primarily a conceptual project, *MRM* at the current stage is designed to illustrate the design utility of scale of intentionality and agency play. Second, as an expressive

project, *MRM* represents our attempt to explore new poetics of computational narratives that depict characters' fluid and subjective internal worlds more than actions [220, 221].

MRM instantiates the design constructs of scale of intentionality and agency play in numerous ways. Ales starts as an avatar with very low levels of narrated intentionality (e.g., agency, emotion, desires, etc.) completely under user's control. As the story progresses, the system dynamically retrieves memories triggered by Ales's encounter of different events, objects, and actors, and computes his belief-system and emotional state. If the recollected memories converge to a coherent personality, Ales's system intentionality increases and he will act autonomously and ignore user's commands contradictory to his belief-system. These memories could also be fragmented, triggering one another and leaving him in a state of confusion and hesitance. The emotional state is used to generate affective descriptions of the world through the computational-cognitive approach to conceptual blending [54, 81].

As we have seen, memory-building is the major means of character formation in the stories generated by *MRM*. Ales's rich mental activities and affective dispositions contrast with his mundane actions. The loose plot structure and emphasis on characters' internal world echo traits of the Modernist stream of consciousness literature, a major inspiration of *MRM*.

Furthermore, *MRM* is an initial step towards a critical commentary on an individual's agency in society. The previous chapter has demonstrated that different approaches lend themselves to distinctive social and political practices. Agency play and the scale of intentionality, hence, seek to offer new means to commentate on a wide range of negotiation and struggle between an individual and her socialization practice, an important component of one's lived experiences. In *MRM*, the power

relation and proportion between user and system agency vary dynamically to highlight a wide spectrum between the human-leader-computer-follower to computer-as-an-autonomous-entity model. Between the two extremes, is what Lewis may call a multi-dominance interaction model between human and computer [129]. Hence, the project can be used as a platform to explore themes such as control, resistance, and dis/empowerment.

MRM is built on a unique interdisciplinary framework that combines AI, cognitive science, and stream-of-consciousness literature tradition. This chapter begins with locating a nexuses of concerns we address. Section 6.1 traces the historical lineage of intersections and divergences between AI, cognitive science, and literature sketched above, highlighting a convergence of concerns that gives *Memory, Reverie Machine* particular salience now. Antecedents to *MRM* in influential and experimental literature and electronic literature are also discussed. Section 6.2 illustrates the ways in which the design principles of “scale of intentionality” and “agency play” are embodied in *MRM*. Finally, Section 6.3 describes the major components of the system, and how they contribute to the overall goal of the system. The following text is based on our presentation “Narrating Artificial Daydreams, Memories, Reveries: Toward Scalable Intentionality in Expressive Artificial Intelligence Practice” at the 2008 Electronic Literature Organization Conference [221] and a paper under the same title which has been submitted for publication.

6.1 *Motivation and Historical Context*

“Stream of consciousness” is a psychological term that William James coined in his 1890 text *The Principles of Psychology* [107]. The term was later applied to works by various modernist writers such as Dorothy Richardson, James Joyce, Virginia Woolf, and William Faulkner, indicating both their literary techniques and the genre itself.

Beyond various formal experiments, stream of consciousness literature reflects a conceptual purpose — to use the internal thoughts as a primary way of depicting fictional characters. As Humphrey puts it in his remarkable study in *Stream of Consciousness in the Modern Novel* [103], the works under this genre replace the motivation and action of the “external man” with the psychic existence and functioning of the “internal man.” A passage from Virginia Woolf’s 1925 novel *Mrs. Dalloway* well captures some key characteristics of this literary tradition:

... so that should she be very old she would still remember and make it jangle again among her memories how she had walked through Regent’s Park on a fine summer’s morning fifty years ago. For she was only nineteen and had got her way at last, to come to London; and now how queer it was, this couple she had asked the way of, and the girl started and jerked her hand...

Decades have passed since the heyday of stream of consciousness literature and many works associated with this literary experiment have entered the canon of “high” literature. Their approach to human subjectivity and focus on pre-speech consciousness nevertheless are still relevant to many recent technologies (e.g., AI), theories (e.g., cognitive linguistics), and forms (e.g., computational narrative). These younger developments, in their own ways, have taken the modernist writers’ steps further in ways described below. Our goal of narrating memory, reverie, and daydreaming computationally in a new form of polymorphic fiction requires an in-depth understanding of the synergy among these fields.

This section first discusses the motivation for choosing computationally generated memories, reveries and daydreams, which will be referred to as the imaginative content, as one of the major themes of *MRM*. Our motivation is situated in the context the synergistic connections between stream-of-consciousness literature, AI and cognitive science research. Section 6.1.2 and 6.1.3 hence are devoted to such connections

and their implications to *MRM*. We believe that the theory, content, and form we present below have collided in this work in a way that is charged at this historical moment where scientists of the mind provide a new lens on modernist literary concerns and where computation provides a new means for engaging both.

6.1.1 Machine Memories, Reveries and Daydreams

Scientific communities have had varying takes on human subjectivity and pre-speech thoughts like memories and daydreaming. Positivist views of science holding sway during much of the modernist period devalued such phenomena as unobservable and therefore unscientific. Later, artificial intelligence researchers, reliant upon a “computer as brain” metaphor, focused on internal thought, but only of the mathematically grounded sort amenable to the language of computation. This perspective was carried through to the early days of the interdisciplinary enterprise of cognitive science. Under such a view, phenomena such as daydreaming must be construed as having a rational utility amenable to modeling. Not surprisingly, the techniques chosen were appropriate to rational, conscious thought processes such as planning [152].

It is no wonder then that speculative fiction on this subject matter has considered human subjective mental phenomena to be the final frontier between AI-powered machines and mankind. Indeed, speculative fiction characters such as the androids David and Gigolo Joe in the film *A.I. Artificial Intelligence* [199], Lieutenant Commander Data in the television series *Star Trek: The Next Generation* [178], and Roy Batty in *Blade Runner* [187] all were constructed to interrogate aspects of human subjective emotion, aspirations, and dreams. This frontier is well depicted by the following dialogue from *Blade Runner*:

Tyrell: Commerce is our goal here at Tyrell. “More human than human” is our motto. Rachael is an experiment, nothing more. We began to recognize in them strange obsession. After all they are emotional inexperienced

with only a few years in which to store up the experiences which you and I take for granted. If we gift them the past we create a cushion or pillow for their emotions and consequently we can control them better.

Deckard: Memories. You're talking about memories.

The potential of generating and narrating the imaginative content (e.g., memories and daydreams) triggered by, and sometimes at odds with the world at hand has not been fully explored by the electronic literature and interactive narrative communities, both of which have been profoundly touched in recent years by advances in computer gaming. Until recently, many gaming oriented works in the fields have focused on advancing the plot through object acquisition, combat, or puzzle solving, leaving behind a full range of psychological aspects of the characters.

Informed by the stream-of-consciousness literature, *MRM* engages such topics through the narration of memories, daydreams, dispositions, and intentions, all of which are used to signify system intentionality. It explores the tension between rigid computational algorithms and fluid human cognitive processes by computationally generating affective depictions of the story world as well as the character's internal world. In summary, *MRM* foregrounds a dance between user and system intentionality through the stories about a robot character who, similar to Lewis' *Voyager*, may possess human-relatable intentionality and independence. The motivation is to bring forth novel narrative experiences as well as critical reflections of social and political conditions, a tradition that has been previously explored by the works of Yoko Ono, Adrian Piper, and Sol LeWitt.

6.1.2 Stream of Consciousness Literature and AI

The stream of consciousness writing and AI may pose an unlikely match as a subject of comparative analysis. The two fields not only sprouted in different historical periods, but also reside in two separate communities. One was populated in the early

twentieth century and is now associated with academic literary analysis more often than being seen as vibrant area for active creative production, whereas the other is an on-going development in the techno-science sphere that underwent significant self-reevaluation after the two “AI Winters” in the 1970s and 1980s [180]. Beneath the obvious differences, however, are the similar overarching goals and historical contexts shared by the two fields.

First, stream of consciousness literature and AI speak to each other through a shared ambition. Humphrey observed that “[t]he attempt to create human consciousness in fiction is a modern attempt to analyze human nature” [103]. If stream of consciousness writers sought their answers by portraying humans directly, the AI community pursued theirs by constructing the “other” — machines. AI practitioner Michael Mateas recently echoed that “AI is a way of exploring what it means to be human by building systems” [136]. These systems, built in attempt to resemble or surpass their human creators, have become our mirrors to reflect upon our identities as humans [203].

Second, both fields rejected behaviorism in their respective historic periods, and turned their attentions to what happens internally in human mental activities as gateways to understanding “human nature.” Prior to the turn of the twentieth century, fictional characters were typically represented by their external behaviors. Writers carefully crafted their actions, dialogues, and rational thoughts to create distinctive personas for their stories. What stream of consciousness writers were able to achieve, in comparison, was to create their characters mainly out of their psychological aspects, including their buzzing random thoughts and associative trails.

The scientific community from which AI grew out of in the 1950s, in parallel, was similarly dominated by behaviorism. The paradigm was based on the laws of stimulus-response and declared itself as the only legitimate scientific inquiry. Mental constructs such as knowledge, beliefs, goals and reasoning steps were dismissed as

unscientific “folk psychology” [180]. Part of AI’s contribution was to bring these scientific taboo back to the table by building powerful computational systems based on them. For instance, the key AI research topics including reasoning, planning, and learning, can mostly be seen as efforts to model the correspondent human cognitive capabilities.

It is also worth pointing out some of the differences between the two areas that are relevant to our project. Although both look at cognitive phenomena, stream of consciousness writers and AI practitioners emphasize different stages of human consciousness. The term “consciousness” from the vantage point of modernist writers referred to “the whole area of mental processes, including especially the pre-speech levels.” This was based on James’ original psychological theory, in which “memories, thoughts, and feelings exist outside the primary consciousness” and, further, that they appeared, not as a chain, but as a stream, a flow [107]. AI, and GOF AI in particular, on the other hand, regarded human rationality as the key to problem solving. Early AI practitioners relied upon the rational and stable operations of our cognitive processes at the cost of the addressing the roles of the body, affect, and the uncontrollable stream of thoughts unmediated by logic and rationality. Such abandoned topics, of course, had been the constant themes under the pens of stream of consciousness writers.

Another difference between AI and stream of consciousness literature is the conflict between the former’s the emphasis on generalizability and the latter’s pursuit for specificity. Modernist writers such as Virginia Woolf firmly believed that an artist’s and writer’s mission was to express her private vision of reality, life and subjectivity. Woolf’s characters in her various novels, such as *Clarissa Dalloway*, *Mrs. Ramsay*, and *Lily Briscoe*, all embodied her belief in an individual’s unique search of meaning and identification [103]. This individualistic approach contrasts strongly with AI’s focus on generalizability, in which individual differences are often sacrificed for regularity

and scalability.

The previous two differences impose challenges to the development of *MRM*. In order to address the first difference, *MRM* attempts to complement both traditions through deploying AI algorithmic works to narrate imaginative (instead of purely logical) cognitive experiences such as of memory, reveries, and daydreams. As for the second, we reconcile these two stances by distancing our work from an attempt to reduce mental activities to uniform formal algorithmic processes. Instead, our project utilizes scientific computational methods, including logical/mathematical formalization, as a way to express our human search for meaning.

6.1.3 Stream of Consciousness Literature and Cognitive Linguistics

The pre-speech level of thought that was neglected by the AI community has been scrutinized again recently in a new field closely built, in part, upon AI: cognitive science. To contemporary cognitive linguists, such as Gilles Fauconnier, George Lakoff, Mark Johnson, and Mark Turner, this neglected land of consciousness holds the basis for our basic conceptual, and even literary, thought [54, 51, 118]. “Language is only the tip of a spectacular cognitive iceberg, and when we engage in any language activity, be it mundane or artistically creative, we draw unconsciously on vast cognitive resources, call up innumerable models and frames, set up multiple connections, coordinate large arrays of information, and engage in creative mappings, transfers, and elaborations,” we are reminded by Fauconnier and Turner [54].

Beneath the tip of this iceberg is what Fauconnier calls “backstage cognition” [52], defined as “the intricate mental work of interpretation and inference that takes place outside of consciousness” [54]. Thus, we could say that cognitive linguists cite phenomena that are even below the unarticulated thought phenomena explored by stream of consciousness authors. Fauconnier sites a range of results in cognitive science to support his conjecture that many cognitive phenomena are rooted in

backstage cognition: “viewpoints and reference points, figure-ground / profile-base / landmark-trajectory organization, metaphorical, analogical, and other mappings, idealized models, framing, construal, mental spaces, counterpart connections, roles, prototypes, metonymy, polysemy, conceptual blending, fictive motion, force dynamics” [52].

It may be argued that one of the reasons that early AI largely confined itself to the territory of rationality is the extreme difficulty that the field ran into in its attempt to model common sense and contextual reasoning explicitly. These powerful, but for the most part invisible, operations are seen within the field of cognitive linguistics to be partially observable in the structure of our linguistic creations. The challenges posed by the cognitive linguistics enterprise offer the opportunity to revisit some of the compromises that AI made in its early stage.

6.1.4 Challenges of Engaging Legacy Forms

In forging this bond between stream of consciousness literature, AI, and cognitive science, we are sensitive to a common trend in computing, that is, trying to prove its triumphs by taking on tasks that seem to be humanly creative in ways that are highly esteemed culturally. Examples include composing music like Mozart [28], playing chess better than a grandmaster [102], or completing complex mathematical proofs [126]. Likewise, we are aware the literature is often viewed in academic disciplines as a succession of movements, tastes, trends, and techniques. Stream of consciousness literature holds a special status in the canon of English literature and therefore imposes unique challenges when used as an inspiration for this computational narrative project. When engaging legacy forms, authors are inevitably put in dialogue with the original works’ historic contexts, social statuses, political agendas, etc. The following clarifies our project’s position in relation to this set of modernist literature works.

Our aim in referencing stream of consciousness writing is not to inherit its high-brow status or to prove the effectiveness of the authors’ system by tackling the “greats” of literature. Nor is it a sign of indifference to more recent modes of literary cultural production. The reason for paying homage to this particular literary form is that it offers critical insights into the AI practice and expressive platform to deploy cognitive science research results. Engaging a legacy form in a different medium requires more than mere translation. The intent of our project is not to generate texts similar to texts by stream of consciousness writers that can “fool” the reader or pass a type of “stream of consciousness writing Turing test.” Instead, we intend to establish a new aesthetic form that calls critical attention to the issues of embodiment, affect, intentionality, and being through internal reflection and imaginative processes.

6.1.5 Related Works

Before we turn fully to our new developments, we call attention to our antecedents. *MRM* speaks to a variety of literary and technical approaches to narratives, such as modernist and avant-garde literature (e.g., stream of consciousness literature, experiments by members of the Oulipo group), algorithmically generated computational narrative systems typically developed in the field of computer science (e.g., Meehan’s *Talespin* [142], Scott Turner’s *MINSTREL* [207], and Bringsjord and Ferrucci’s *BRUTUS* [19]), and the broader domain of interactive narrative (e.g., interactive fiction works, and video games).

The discussion at this section narrows down to several computational and interactive narrative systems that are specifically related to *MRM* either at the content (e.g., daydreams and reveries) or the discourse level (e.g., system agency and intentionality). Two computer science based systems are discussed first with their close focus on daydreaming. Next, we will examine two interactive narrative works especially

regarding their use of system agency.

6.1.5.1 DAYDREAMER

Erick T. Mueller’s program *DAYDREAMER* [152] bears a close resemblance to *MRM* as both systems algorithmically generate narratives that can be read as daydreams and reveries. *DAYDREAMER* takes as input the description of a situation, such as an accidental encounter of a movie star, and produces daydreams “spontaneously recalling or imagining personal or vicarious experiences in the past or future” [153]. Mueller and Dyer provided several daydreams which their theory of computational daydream intends to model. Two sample daydreams based on empirical data are included below with their initial situation:

Last night I went alone to the Nuart Theater to see a film. This actress whom I’ve always had a crush on also came alone and happened to sit down near me. I recognized her and started a conversation. At one point, she mentioned the obviously well-known director of the film we were about to see and I didn’t know the name. I was embarrassed. I finally asked her if she’d like to go have a drink after the film. She said she had other plans.

Daydream 1:

I feel embarrassed for not knowing who the director was. I should’ve pretended to know who the director was. In the future I’ll read up on the film before I go.

Daydream 2:

I’m disappointed that she didn’t want to go out with me. I imagine that she accepted my offer and we soon become a pair. When she has to do a film in France, I leave my job and travel there with her. As we continue

to travel, my career suffers. I become unhappy and feel unfulfilled. She loses interest in me, because I have nothing to offer her. It's good that I didn't get involved with her, because it would've led to disaster. I feel less disappointed that she turned me down.

These samples underline daydreams' cognitive and psychological utilities, namely, rationalization, revenge, failure/success reversal, and preparation. *DAYDREAMER* is essentially a planning system that generates plot units achieve a certain control goal specified both by an emotion state and a personality model. For instance, a plan for rationalization is employed in *Daydream 2* in order to achieve the rationalization control goal activated by the emotion of disappointment resulting from being rejected by the actress.

The planning-based approach, on the other hand, is at odds with *MRM*'s aesthetic and conceptual goals. Widely as it is adopted in computational narrative research [177], the planning-based approach leaves a strong impact over the narratives it produces. From the first computational narrative system implemented with a planning algorithm, *Tale-spin* [142], stories generated along this line often have a noticeably goal-driven and action-based flavor, echoing their underlying algorithmic processes. This aesthetics imposes an antithesis to *MRM*'s aim of generating fluid and loosely structured narratives, preceded in stream-of-consciousness literature. In *MRM*, daydreams, memories, and reveries are not goal oriented, but rather reflect the protagonist's subjective and floating internal thoughts.

Furthermore, the goal-driven behavior entailed by planning-based system is culturally associated with the cold "machineness." The AI cultural icons, such as DATA from TV series *Star Trek: The Next Generation*, or HAL from Kubrick's *2001: A Space Odyssey*, all unmistakably display traces of planning-based behavior (e.g., playing chess [137]), which is used to indicate their differences from, if not inferiority to, human. The concept of system intentionality, however, intends to complicate the

simplistic distinction between human and machine. Avoiding the usage of planning in *MRM* therefore is not only to explore new narrative styles, but also to limit the cultural connotation associated with it.

6.1.5.2 Visual-Daydreamer

Pérez y Pérez et al’s recent program *Visual-Daydreamer* [170] provides an alternative approach to daydreaming generation. Based on Pérez y Pérez’s engagement-reflection model [167], daydreaming is implemented as the engagement or generation phase, that is, the current state of the story is used as a cue to retrieve possible actions from the library of pre-authored story snippets. In other words, the essence of daydreaming here is not the narrative itself, but the open-ended process of generating it. *MRM*, in comparison, focuses both the underlying generative mechanism and the presentation of the daydream and memory components in the narrative.

Unlike the standard planning systems such as *DAYDREAMER*, which generate a sequence actions that connect an initial state and a predefined desired end state, *Visual-Daydreamer* only contains a start state and hence produces unpredictable and loosely coherent stories that are reminiscent of daydreams. A daydream example provided by the authors is below:

A attacks B
B fights with A
A wounds B
A runs away
B dies

MRM shares *Visual-Daydreamer*’s approach to construct daydreams based on similarity. In the case of the latter, similarity is drawn at the story context level — which

existing story in the library has similar context as the current state of the story. Currently, *MRM* retrieves memories and daydreams based on similarity of story content (e.g., themes, topics, ideas, and objects) through annotated keywords. It is our plan to further explore daydream and memory retrieval using analogy-based approaches.

6.1.5.3 Violet

Jeremy Freese's *Violet*, the first place winner of the 14th annual Interactive Fiction Competition, offers a good example of the use of system intentionality as the main theme of narrative. Its plot line is that the protagonist, an easily distracted graduate student, needs to complete 1,000 words in his dissertation in order to win back his girlfriend, Violet, who is frustrated by his slow progress. Compared with other well-acclaimed IF works such as Adam Cadre's *Photopia* (1998) and Emily Short's *Galatea* (2001), the goal of *Violet* seems straightforward and clearly stated from the very beginning. However, the way to achieve this goal is no less puzzling than the earlier IF-based adventure games — every time the user issues the command “write,” obstacles or distractions always occur and prevent the protagonist from doing so. The challenge of the user is to figure out ways to resolve the obstacles and block distractions so that the protagonist can move ahead in his writing, as shown from the below excerpt (the story is narrated using the voice of Violet):

> write

You're standing. You've been pacing most of the morning, stewing about our fight last night.

I don't myself see why one can't type just as easily standing up as sitting down, but you are adamant that you can only focus enough to write seriously if you are seated.

> sit down

Okay. You are seated at your desk and ready to go. Just write!

> write

Unfortunately, with all the pacing about and ruminating, you've been here over two hours but haven't actually opened your word processor yet.

> open word processor

As you move your hand to open the word processor, you look briefly at the tattoo on the back of your hand.

Open. Chapter 3 of your dissertation awaits. You can do it!

> write

You start trying to focus on the screen and type, but you are very tired. I realise that you hardly slept last night between being all worked up from our fight and then being banished to the dodgy futon in our living room.

Still. I know you won't let a little grogginess stop you. You are determined. For us! You look again at the tattoo on the back of your hand for inspiration.

The battle between the system's intentionality, manifested as the protagonist's desire to pursue his distractions, and the user's intentionality to make him write the dissertation, among other features of *Violet*, distinguishes it from similar IF works. Far from simply a formal experiment of a semi-autonomous player character (Section 5.3.4), the split of control over the protagonist between the system and the user plays a narrative role. It represents the two conflicting aspirations simultaneously present in the protagonist's character.

6.1.5.4 Lost One

The presence of both user and system intentionality is a step towards a scale of intentionality. Next, a mechanism that can modulate system intentionality is needed. Although not specifically geared towards modulating intentionality and agency, Nick

Montfort’s work [149] on computationally manipulate narrative discourse is of particular relevance. Built upon Genette’s seminal work in narratology [68], his work developed a formal theory of narrative variations in the context of IF, natural language generation, and computational linguistics.

Montfort’s IF work *Lost One* [149, pp.125-29] is an interactive narrative piece whose discourse can be adjusted to demonstrate a varying level of narrative distance, that is, how immediate a story is narrated. It is of particular relevance to *MRM* because both works seek to gain finer grained control of narrative at the discourse level, regardless whether the end goal is narrative distance or system intentionality/agency. This scalability is very different from *Violet*, where the level of system intentionality remains unchangeable.

In *Lost One*, the narrative distance is represented along a number of dimensions. In order to increase the distance, for instance, the system may 1) insert distancing phrases (e.g., “It appeared that,” and “It was as if”), 2) switch the time of narrating from simultaneous with to subsequent to the events (i.e., from the present to the past tense), 3) reorder the chronicle order of events to other less immediate temporary order (e.g., retrograde, and random), and 4) switch from second person (“you”) narration to third person (“the visitor”). Some of these methods are closely related to *MRM*. As shown in Section 6.2.1, the perspective from which the story is narrated plays an important role in suggesting the scale of intentionality in *MRM* — a story narrated from the protagonist’s personal and affective perspective implies higher system intentionality than one told from an objective narrator’s point of view.

The following two excerpts demonstrate the difference in their narrative distances, achieved mainly through the use of random temporary order, and third person narration:

Example 1:

You look at the center of the plaza.

Your senses are humming as you view the broad, circular, encircling Plaza of the Americas. The morning has concluded. It is midday now.

From here, you see in the distance a statue and a flaneur to the north, a fountain to the east, a trash collector to the southwest, a ball and a boy to the northeast, an obelisk to the south, and some punk and a tree to the west.

The flaneur goes to the northern area.

The punk kicks the tree.

The boy throws the ball.

The trash collector picks up something.

The ball falls to the ground.

Example 2 (increased narrative distance):

The flaneur went to the eastern area, perhaps.

Then, the punk conveyed herself to the southern area, if memory serves.

Then, the visitor looked at the center of the plaza.

Her senses were humming as the visitor viewed the broad, circular, encircling Plaza of the Americas. No doubt, the morning had concluded. It was midday then.

From there, the visitor saw in the distance the statue to the north, the flaneur, the trash collector, and the fountain to the east, the trash collector to the southwest, the boy to the northeast, the obelisk and the punk to the south, and the punk and the tree to the west.

Then, the trash collector took something.

Before that, the boy took the ball, or so it seemed.

Before that, the visitor walked to the center of the plaza.

Lost One also engages expressive use of system agency (and hence system intentionality). The overall story is that the protagonist is waiting for a friend in a plaza. The user can choose to stay put, which will remain the narration at its low narrative distance. If the user, on the other hand, directs the protagonist to move around in the plaza, the narrative distance increases, indicated by the above mentioned methods. At any point the user can stop to stay put, which will decrease the narrative distance gradually. However, if the user keeps moving, the system will suddenly eliminate her agency in the story and leave the protagonist wandering about at random under its own control. Compared to *MRM*, the balance between system and user agency here is reversed all at once. In terms of scale of intentionality, *Lost One* represents a binary case where the agency either resides completely in the user or in the system, whereas *MRM* intends to provide a finer-grained scale between these two extremes, paralleling *Lost One*'s goal to modulate narrative distance.

6.2 *Deployment of “Scale of Intentionality” and “Agency Play”*

MRM is designed specifically to explore the new narrative possibilities brought forth by scale of intentionality and agency play, both of which were described in Chapter 5. In a nutshell, *MRM* intends to construct various levels of system intentionality through independently adjusting the following aspects of its main character Ales: 1) the amount of user's control of Ales, 2) the amount of the imaginative content (e.g., memories, reveries and daydreams Ales retrieves), and 3) the perspective from which the stories are narrated. This section discusses the specific ways in which these design principles are incorporated in the system, and illustrates the particular uses through three examples at the end of this section.

6.2.1 Main Indicators of System Intentionality

MRM currently deploys system intentionality at the following three levels:

Action level The design principle of agency play is mainly implemented at this level. The actions of Ales are overall jointly controlled by both the user and the system. When emphasizing user’s agency, *MRM* directs Ales mainly following the command input from the user. On the other hand, the system will make autonomous decisions about the main character’s actions, should system agency needs to be highlighted. In the intermediate level, Ales may express his different opinions, computationally derived based his beliefs and desires, but still reluctantly follows user’s command. Ales is an example of what is previously (Section 5.3.4) defined as the semi-autonomous player characters (SPC), over which the user’s and the system’s agency share an inversely dependent relationship. It is straightforward that the more Ales’s actions are controlled by the system, the higher the system agency and hence higher system intentionality.

Content Level At the content level, the scale of intentionality is externalized by the ratio of the amount of main story content to the amount of memories, reveries and daydreams. Compared to the main story in which the user may decide a substantial portion of the main character’s actions, the imaginative content is beyond the direct control of the user at all times. As Section 6.3 will explain in more depth, a piece of memory or daydream is typically evoked automatically by certain events, objects, and actors in the narrative, if the current system intentionality surpasses a certain level. Although the user may impact the artifact or other characters that Ales interactive with, which in turn triggers certain imaginative contents, she does not have direct access to which memories and daydreams that the system will retrieve. In other words, the process of how certain imaginative contents are generated remains to be somewhat opaque to the user. As the system evokes more imaginative content, the system intentionality of the main character therefore increases accordingly.

Perspective Level As stated earlier, Montfort [149] illustrates the importance of narrative discourse as a narrative dimension, independent from the narrative’s content from the vantage point of interactive fiction works. *MRM* echoes Montfort’s observation by utilizing the narrative discourse, particularly the perspective from which the story is narrated, as a third indicator of system intentionality. In *MRM*, the same narrative content can be told either from a third person’s perspective with straightforward and descriptive tones, or be narrated from the Ales’s perspective signified with more subjective descriptions and affective dispositions, generated dynamically by Harrell’s Alloy conceptual blending algorithm (details in Section 2.3.3.2). The extent to which the system narrates stories from the perspective of Ales, which suggests the main character as an independent character with beliefs, desires and other intentional mental states, determines the level of system intentional.

Table 10: Indicators of “Scale of Intentionality” in *MRM*

Level	The ratio of
Action level	Autonomous actions / user selected actions
Content level	Imaginative content / the main narrative
Perspective level	Ales’s perspective / third person perspective

6.2.2 Examples of Various Levels of System Intentionality

The above indicators of the scale of intentionality is summarized in Table 6.2.1. The higher the ratio at each level, the more system intentionality *MRM* intends to achieve. Ideally, these levels can be manipulated independently to achieve various effect; however, the current version of *MRM* adjust all three levels simultaneously. This means when a higher level of system intentionality is needed, for example, *MRM* increases the ratio at all three levels altogether. In order to discuss the engagement of scale of intentionality concretely, the rest of this section provides three short examples of a very similar scenario generated by *MRM*. Each example includes one iteration of interaction between the user and the system, and represents a different level on the

scale of intentionality. Notice a user's comment is represented in capitalized words.

- Example One: high user intentionality

(ales approached the door. in his visual scan it was painted [red, yellow, blue, grey, or brown?] ...)

RED

(.)

(he raised his mechanical arm to [knock, open, open gently, punch, or open carelessly?] the door ...)

- Example Two: high system intentionality

(at the end of a long corridor a door loomed. ales looked up; he decided to see it as (red, yellow, blue, grey, or brown?) ...)

(blue)

(. it was dusty wood , almost exactly like the one in his mother's bedroom.)

(the room where he had his first encounter of tune-up and oil change had similar doors.)

(the oil change left a sickly feeling in his gut)

(he would rust like the tin man before enduring another.)

(he raised his steel arm to [knock, open, open gently, punch, or open carelessly?] the door...)

- Example Three: joint user and system intentionality

(ales stopped, computed, and moved again to (knock, open, open gently, punch, or open carelessly?) the door ...)

PUNCH

(The door creaked into an entrance.)

(the funny noise reminds him of)

(the hospitals, the junkyards, he went to frequently)

(, which was so unpleasant that he always avoided it ever since. ohhh, the same noise.)

(he paused a bit before he walked in.)

(the bothersome light-colored room looked somewhat like the room he was born.)

(his disturbingly familiar mother probably had left him quite early.)

(most of his knowledge about her came more from)

(the engineer pointing to the women in pictures)

(his memories of experiences.)

(the things she liked to wear, the way she smiled.)

6.2.2.1 *Action Level*

In short examples provided above, an immediately noticeable difference is that in Example 1 and 3 the user is able to determine Ales’s action, whereas the system acts autonomously in Example 2. As stated earlier, the system’s capacity to act autonomously is one important indicator of a high level of system intentionality, and particularly system agency. The rest of the discussion at the action level centers around different layers of agency (Section 5.3) in *MRM*.

The scope of impact of user actions, besides their immediate consequences and longer-range effects on the main story world, can also manifest itself through the emotional tone of the blends and selection of particular memories episodes, both determined by the emotional state of Ales. In Example 3, the user directs action to “punch” the door open does not only allow the main character to enter the room (local agency), but also inform later interactions related to Ales’s experience in the room (global) via a history of actions. Furthermore, it also moves Ales’s emotional state into “anger” (currently implemented using a simple finite state machine, described in 6.3), which will influence the emotional tone of blends and memories in both immediately

and in the longer term future. A major distinction between Example 2 and 3 is that in high level of system intentionality (Example 2), system action is selected based on Ales’s existing emotional state. In Example 2, Ales happens to be in a “gloomy” mode, which leads to the choice of “blue” due to the latter’s similar emotional tone. Once the system reaches very high level of intentionality, *MRM* will autonomously choose actions that fits Ales’s emotional state. In a moderate high level, Ales could refuse to carry on a command issued by the user that is contradiction his emotional state, or obey the command “reluctantly.”

Although the system is capable of generating highly variable output, certain moments in the stories are designed to have low global agency in order to achieve expressive goals. For instance, in another short narrative sequence of the output, the protagonist SPC is a cleaning robot working in a kitchen. The beginning of the story offers a great variety of tasks to Ales, and user choice of action results in detailed description of the consequence. As the story continues, and the character gains more domestic work experience, both the number of options and the level of details decrease. In one particular moment, the user has to repeatedly choose between “wiping” a dish and “rubbing” a dish, two closely related actions, yet the outcome remains unchangeable. The aesthetic choice of emphasizing (limited) local agency over (even more) limited global agency in this particular moment of the narrative is set to contrast with previously encountered narrative sequences that incorporated a higher magnitude of user agency relative to system agency. This design decision therefore conveys the emotional state of the character: boredom with a repetitive daily job, in addition to a sense of confinement and powerlessness. The sundry moment is meant to evoke a sense of pathos and meaninglessness while completing a never ending list of sundry chores.

Different dimensions of agency play are guided by user input direction, i.e. they are mediated by patterns of user input. A repeated series of user choices reinforces the

main character’s emotional state and patterns of behavior thus impact the system’s style of narration, the character’s action performed, the short and long term impact of actions, and dynamic changes in the degree of user control versus system control.

6.2.2.2 Content Level

As described earlier in this section, the imaginative content including Ales’s memories, reveries and daydreams is determined by the system, triggered by objects that appear in the narrative, and is not directly controllable to the user. *MRM* generates such imaginative content indirectly based on user’s interaction, such as the artifacts/characters that she chooses to interact with, and/or dispositions with which actions are performed (e.g., “punch” vs. “open gently”). At a low level of system intentionality as in Example 1, the amount of imaginative content is very limited. At a high level, on the other extreme, cascades of memories and daydreams may trigger one another and completely take over the main story line.

6.2.2.3 Perspective Level

Examples 1 and 2 also differentiate themselves in terms of the narrative perspective. It is true that at the first glance both texts were written from the third person perspective using the pronoun “he” to refer to Ales. However, a closer examination shows the difference. The narrative in Example 1 is told from an “objective” third person point of view that emphasizes actions, facts, and consequences, whereas Example 2 is comprised of “subjective” memories and “affective” description based on the “perception” of Ales. In fact, the decision not to use direct first person pronoun of “I” is to maintain a stylish connection to Woolf’s stream of consciousness works, which characteristically combines third person voice with highly affective first person experiences. Readers who are familiar with interactive fiction convention may notice that the typical second person (“you”) is also avoided, a decision made for the above reason.

The goal of narrating the story from a first person perspective (different from first person voice) at a relatively high level of intentionality is mainly achieved through blends. In *MRM*, blended descriptions are created from mixing properties of “objective” elements (e.g., events, objects, and actors [205, 78]) and affective dispositions (e.g., Ales’s current emotional state) algorithmically determined by the system. For instance in Example 3, logical axioms selected from an ontology describing the concept “mother” are blended with axioms describing affective concept “anger,” into a “disturbingly familiar” mother. It is used to highlight that the story world is what Ales “*perceives*” it as, influenced by his mental states, emotions, etc. It is important to be clear that blending here is not the mere concatenation of words to form compound phrases. It refers to the conceptual integration of multiple concepts according to a set of cognitively-inspired optimality principles (details refer to Section 2.3.3.2). In this case, modifier-noun phrases, some of the simplest indicators of conceptual blends, are the final result of an underlying process that is semantic, not lexical.

Individually speaking, Example 3 offers the most interesting effects, for it provides a scenario where the user and system’s intentionality and agency are conflated. However, the combination of all three examples provides an over-arching theme of the story, that is, the (re)gain of identity of the main character. The contrast between Ales’s low intentionality in the beginning of the story and his high intentionality at the end suggests the transformation of the character. From the user’s perspective, the unconventional experience of slowly losing control of a character is intended to evoke critical reflections of the theme of control and hierarchy at various levels.

6.3 Major Components

So far, this chapter has situated the work in a historical web where stream of consciousness literature, artificial intelligence discourse, and cognitive science research complement each other. This correlation converges with the authors’ aesthetic goal

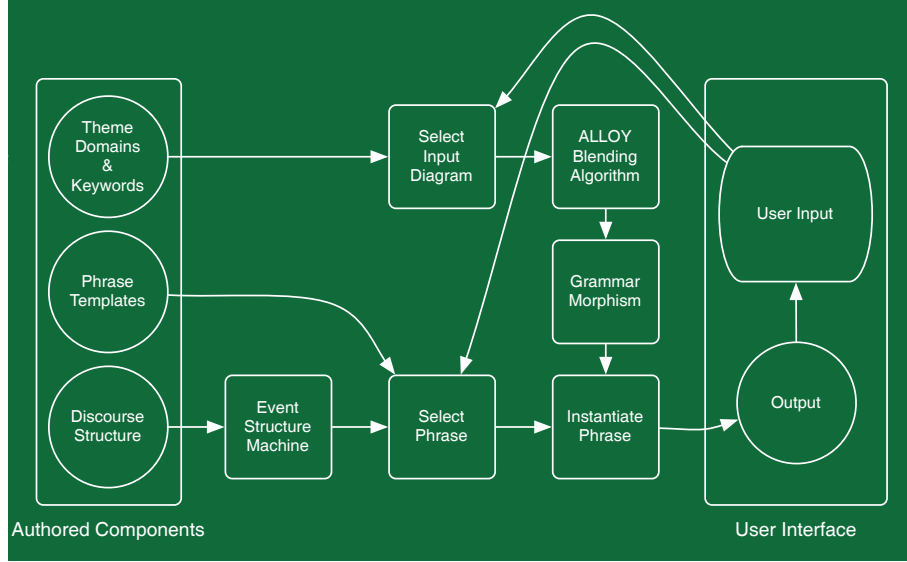


Figure 26: Architecture Diagram of *GRIOT* [77]

and calls for a new form of computational narrative, on the subject of daydreams, memories, and reveries. We have also discussed the various ways in which the novel design concepts of scale of intentionality and agency play are embodied in *MRM*. This section provide a high level overview of the major components that are essential to the deployment of scale of intentionality. The explanation is illustrated with Example 3 from the previous section, where different elements in the output are annotated in Figure 27.

The GRIOT system [77] (Section 2.4.2) is the foundation of *Memory, Reverie Machine* both in terms of technical implementation and our approach to computational narrative [80]. A more detailed technical description can be found in Harrell’s dissertation work [81], where the detail of Figure 26 is explained.

Before we start, it is important to distinguish between the project’s various levels of technological and expressive investigation and production: (1) the system as an abstract model for how computational narratives can be made generative, extensible, and reconfigurable, (2) the system that generates the story, (3) the narration techniques developed, (4) the story content, and (5) each instance of output.

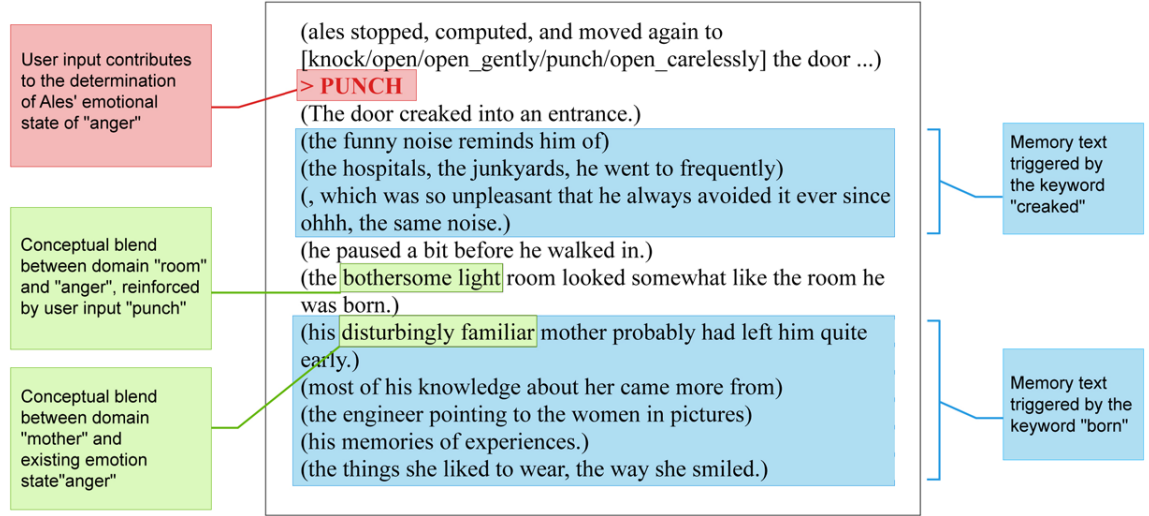


Figure 27: An Annotated Example of *MRM*'s Output

The emphases in this section are upon (1) and (2), which comprise our technological framework as inspired by the historical context and theoretical and creative frameworks above, and secondarily on (3), the narrative techniques to depict inner thoughts, and (4) the self-reflexive subject matter. (3) is heavily influenced by Virginia Woolf's stream of consciousness novel *Mrs. Dalloway* [218]. Below we highlight particular aspects of the system relevant to the technical and expressive goals involving the narration of system intentionality.

6.3.1 Dynamic Narration of Affect Using the Alloy Conceptual Blending Algorithm

Computationally, our system builds upon the GRIOT framework, which identifies, formalizes, and implements an algorithm for structural aspects of conceptual blending with applications to computational narrative. Alloy is the primary generative component of the GRIOT system. The Alloy algorithm, when modestly applied in *MRM*, generates blends involving connecting current experiences of events, objects, and actors [205, 78] to affective concepts determined by current state, in this case emotional state, of the protagonist (discussed below). In the example in Figure 27,

logical axioms selected from an ontology (semantically structured database) describing the concept “door” are blended with axioms describing affective concept “anger.”

The Alloy algorithm uses a set of formal optimality criteria to determine the most common-sense manner in which the concepts should be integrated [81]. The result is a blended axiom or set of axioms that is then mapped to natural language output. For example, the description of the door ranges from “distasteful wood-colored” to “irritatingly sturdy” or more depending on the concepts being blended. Since blending refers to the conceptual integration of multiple concepts, it is important to be clear that blending is not the mere concatenation of words to form compound phrases. In this case, compound phrases, some of the simplest indicators of conceptual blends, are the final result of an underlying process that is semantic, not lexical. It is not difficult to imagine that the phrase “inviting gateway” from the sample output in introduction comes from blending concepts of “door” and “happiness” and trying to articulate the outcome in the English language.

Constructing blends between objective and affective concepts allows us to achieve a balance between author determined plot and variable theme or emotional tenor. An artifact required by the plot can be depicted in various ways based on the character Ales internal emotional state. The highly subjective description, in turn, portrays personality traits of the character, a recurrent technique in *Mrs. Dalloway*.

6.3.2 The Emotional State Machine

Actions taken by a character in a computational narrative, which are usually (but not exclusively) selected by a user, can guide building up of a profile that describes user’s preferences, history of actions, and analysis of trends in those actions. A simple, but effective, step in this direction is tracking the emotional state of a character based upon actions that the character has taken. *Memory, Reverie Machine* allows user to directly influence the emotional state of Ales, and hence the selection of affective

concepts for blends. She may choose among an array of pre-defined actions, such as seeing objects as “red,” “yellow,” “blue,” or another color in the robot character’s optic sensors, each connecting to a particular emotion. A keyword “red,” for instance, may trigger an affective concept “anger.” These emotional mappings are designed aesthetically by the authors to achieve narrative effects, not as an attempt at cognitively modeling emotion using a computer as has been the stated goal in multiple traditional AI projects.

A successful interactive narrative requires a careful balance between the user’s agency and author’s intention. In our system, user’s impact on the character’s emotion is moderated by the emotional state machine component for the sake of narrative consistency. The state machine records Ales’s current emotion based on the entire history of user input, instead of the most immediate one. It guarantees that changes of Ales’s emotions will be gradual, even if user input oscillates between opposite emotions.

6.3.3 Memory Structuring and Retrieval

In *MRM*, we seek to make output coherently extensible at runtime. For this project we allow the narrative to be punctuated with episodic remembered events and longer reveries of remembered experience. Again, this is not meant as an experiment in cognitive modeling or advanced algorithmic design, our goal is to demonstrate discourse that is meaningfully reconfigurable to serve an author’s expressive goals in dialogue with a user’s selected actions. In this case, each memory is annotated based on its subject matter and is retrieved when at least one subject item appears in the story line. In the example in Figure 27, Ales’s unpleasant memory of hospitals and junkyards is triggered by the opening of a door through the mutual subject of a certain sound. The system also keeps track of the emotional tone of each memory and selects a memory only if it does not clash with the current emotional state. The example at

the start of the Introduction also illustrates this feature.

We have described the major components of *Memory, Reverie Machine* that align with our tale of inner-thought and AI in historical concern through illustrative examples. To result in a longer form interactive and generative story, of course, requires expanding each component. In the next section we reflect upon our accomplishments and challenges and indicate both near and long term future directions for the project.

In summary, *Memory, Reverie Machine* is an interactive narrative system designed explicitly to reflect the design principles of scale of intentionality and agency play. It highlights a dance between user and system intentionality through the narration of memories, intentions and dispositions. This chapter has described the authors' motivation of algorithmically evoking memories and daydreams, and how this goal is consistent with an interdisciplinary framework that combines stream of consciousness literature, AI, and cognitive linguistic research. Detailed account of how system intentionality and agency are instantiated and adjusted for narrative purposes composes a major part of our discussion. Particularly, we have examined the narrative at the action, content, and perspective level. Finally, major components of *MRM* was described.

Conceived as a critical technical practice [3], and acknowledging cognitive linguistics critiques of computational approaches to cognitive modeling [49], this work is not an attempt to reduce mental activities to a formal algorithmic process. We are inspired by the fact that algorithmic and knowledge engineering approaches themselves can be expressive as in the practice of “expressive AI” [135] and its recent generalization into “expressive processing” [209].

CHAPTER VII

CONCLUSION AND FUTURE WORK

A “program” which could produce music as [Chopin or Bach] did would have to wander around the world on its own, fighting its way through the maze of life and feeling every moment of it. It would have to understand the joy and loneliness of a chilly night wind, the longing for a cherished hand, the inaccessibility of a distant town, the heartbreak and regeneration after a human death. It would have to know resignation and world-weariness, grief and despair, determination and victory, piety and awe... Therein, and therein only, lie the sources of meaning in music.

— Hofstadter, D. *Gödel, Escher, Bach* [93], 1979

Can computers have intentionality? This is an underlying question central to this dissertation. On the one hand, our knowledge about the fundamental differences between humans and computers, as well as our lived experiences, all points to a negative answer. As Hofstadter articulated beautifully above, the depth of human emotions and our socialization into the everyday practices of our culture, the basis of intentionality, is not replicable by computer in the foreseeable future, if ever. This is the position many theorists take, including Searle. On the other hand, we have witnessed an increasing cultural importance of computer systems that are designed to appear intentional, to perform tasks, and to exhibit complex behaviors usually seen as the territory of intentional human phenomena. These systems vary from household appliances (e.g., Roomba vacuum robot) and virtual characters (e.g., Sims characters) to expressive artifacts such as *Voyager* and *EMI*. Compared to programs such as *Adobe*

Photoshop, these intentional systems seem to produce output “about” something in the world rather than the merely execute algorithmic rules. Even Hofstadter admits, upon playing on the piano a mazurka composed by *EMI*, “I was impressed, for the piece seemed to ‘express’ something. If I had been told it had been written by a human, I would have had no doubts about its expressiveness. It sounded slightly nostalgic, had a bit of Polish feeling in it, and it did not seem in any way plagiarized” [96].

The contradiction between the ontological contemplation and our actual interaction with certain computer systems is the driving force behind the theoretical inquiry and design space exploration of intentional systems. This chapter revisits the major arguments made in the previous chapters. After discussing the major contributions of the study, it concludes the dissertation with future research directions.

7.1 Revisiting the Major Arguments

In the discourse of AI, system intentionality is typically seen as a technical and ontological property of the program and is a result of its underlying algorithms and knowledge engineering. Drawing from the areas of hermeneutics, actor-network theory, cognitive semantics theory and philosophy of mind, this dissertation proposes a humanistic view of system intentionality and AI. Specifically, this dissertation explores two central research questions: 1) How is system intentionality formed? and 2) What are the design implications for building systems that utilize such intentionality as an expressive resource?

7.1.1 The Formation of System Intentionality

The primary subject of the study is *intentional systems*, defined with four criteria: 1) operational opacity, 2) system autonomy, 3) human-relatable behaviors, and 4) authorial intention. The term “intentionality,” following a prototype model, encompasses not only AI systems but also AI-like systems that exist either outside of computer

science communities or that are not described by their authors as AI systems for ideological or other reasons. The social and cultural dimensions of the four criteria call for not only a new perspective on system intentionality that does not merely consider the underlying machine operation, but also a novel framework that puts computer systems back into the social contexts where they are created and used.

This dissertation proposes that system intentionality arises from a complex meaning-making network that incorporates system authors' discursive narration and users' hermeneutic interpretation of system intentionality in a broad social context — the *AI hermeneutic network*. In other words, intentional systems are seen as texts because they facilitate the exchange of meaning between their authors and users. Informed by Agre's [3] observation of the importance of intentional vocabulary in AI practice and Hayles' [87] critical work on Alife, this dissertation focuses on the system authors' discursive strategies in narrating system intentionality both at the source code (functions, algorithms, and system architecture) and at the representation level. Different from many AI practitioners such as Agre and McDermott, this dissertation argues that intentional narration is not an epidemic that needs to be avoided, but is rather a constitutive component of the practice of AI. Finally, this framework stresses that system intentionality needs to be understood within social contexts along with other actants (e.g., other relevant systems, funding agencies, and mass media).

The application and value of the AI hermeneutic network construct is illustrated through a close reading of an intentional system, *Copycat*. The analysis of the system's source code and a substantial corpus of technical literature published by the system's authors illustrate their various discursive strategies to narrate the system's technical operation as intentional. Based on empirical data, this dissertation demonstrates the ways in which intentional vocabulary serves as a connection point that links the authors' technical and intentional narrations, both of which are necessary for AI.

7.1.2 Design Implications

Since system intentionality is discursively narrated and interpreted in its social settings, instead of being merely understood as an intrinsic and static property of the algorithmic complexity, it may be manipulated for expressive purposes. In this regard, this dissertation proposes a new design principle called a *scale of intentionality*. It allows system intentionality to be dynamically adjusted along a scale. Similar to the use of a camera, which can be one effective way of storytelling in films, this design approach turns system intentionality into an expressive mechanism that can be used in different forms of digital artifacts.

One aspect of a scale of intentionality (and vice versa), *agency play* [84], is discussed in depth. Informed by the various approaches to a human individual's agency in social theories, this dissertation critiques the conventional view of "agency as free will" in digital media, often followed by the motto of "the more agency the better." By breaking down agency into its various dimensions (agency relationship, scope, dynamics, and user input direction), this new model of agency play offers specific ways to manipulate each dimension of agency as a mechanism to convey meaning and express ideas. In the domain of digital media, it allows a more nuanced understanding of agency that is in dialogue with social theories of this concept.

The new design principles of scale of intentionality and agency play are illustrated in my joint interactive narrative project *Memory, Reverie Machine (MRM)*, co-developed with Fox Harrell. Different from many other works in the same genre, the main character in *MRM* varies dynamically along a continuous scale between a user-controlled avatar with low intentionality and an autonomous non-player character with high intentionality. Using *GRIOT*'s facility for algorithmically controlling the semantic hooks for interpreting system behavior as intentional in the narrative discourse, the authors of the system turn system intentionality into a scalable expressive dimension in interactive storytelling. Expressively, the content and aesthetics

of *MRM* are inspired by the fluid depiction of human subjectivity in stream of consciousness literature. Technically, this work extends Harrell’s *GRIOT* system [81].

Overall, this dissertation represents my approach to combining critical theory and expressive cultural productions, and to blurring the boundary between what are often perceived as an objective, clean scientific practice and subjective, messy human activities. The theoretical framework with the case study of *Copycat* is intended to highlight the “humanness” in the scientific/engineering practice of AI by foregrounding the authors’ discursive role as the spokespersons of their systems. The expressive exploration of *MRM*, in the meantime, can be seen as an attempt to depict the human subjective and affective experience of memories and daydreams with formal and algorithmic machine operations. Similar to my approach to system intentionality, my objective here is not to argue about the fundamental differences between human and computer, or to suggest that one should take over the other. The shift of social discourse from order and homogeneity to multi-dominance and diversity has rendered these discussions less relevant to our contemporary culture. What drives my work is the new possibilities of emergent meaning and reflection on our human condition as a result of the tension and synergy between human and computer.

7.2 *Contributions*

This section discusses the major contributions represented in this dissertation. It is presented in an order parallel to the structure of the dissertation.

AI Hermeneutic Network This construct provides a new framework with which to study the intentionality of computer systems, a phenomenon that has been studied mainly with computers in isolation from their creators, users, and contexts. The AI hermeneutic network is a unique contribution in the following ways:

- A New Perspective on System Intentionality

The topic of intentionality in the context of computers has been discussed since the early days of AI. Most of the debates between philosophers and AI practitioners have occurred at the philosophical and ontological level, based purely on the technical operations of computers. Certain philosophers [35] theorized system intentionality as a derived phenomenon, perceived by humans. According to this theory, our tendency to read intentionality was seen mainly as a survival instinct. However, this approach leaves almost all artifacts subject to similar intentional reading without addressing why certain artifacts seem more intentional than others. This dissertation proposes a new perspective on system intentionality, one that emphasizes its narration and interpretation by human authors and users, who communicate meanings through intentional systems in their social and cultural context. In other words, the argument presented in this dissertation is not a philosophical one, but a social and cultural one.

- **An Analysis of Discursive Strategies of AI Practitioners**

Within the traditional value system of AI, the contributions of its practitioners are often measured in terms of the technical properties of their systems (e.g., algorithmic complexity, knowledge representation, and performance). In reality, the reception of their systems is also affected by how “intentional” and “intelligent” they seem. Different from the AI practitioners who argue that the use of intentional vocabulary needs to be minimized, this dissertation argues and demonstrates with examples that narrating the system intentionality of machine operations is a necessary component of AI practitioners’ practice. Their various discursive strategies need to be acknowledged and recognized, rather than downplayed by the community.

- **An Interdisciplinary Theoretical Framework**

This dissertation has presented an interdisciplinary theoretical framework, drawn

from a broad range of theories including hermeneutics, philosophy of mind, actor-network theory, and cognitive semantics theory. These different approaches are synthesized in a unique way that bridges the divide between humanistic interpretations and algorithmic operations. Applying hermeneutical interpretation to AI code and literature illustrates an especially new use of hermeneutics and a new method for understanding computational systems.

A Close Reading of a Notable AI system In order to illustrate the utility of the AI hermeneutic network for analysis and interpretation of intentional systems, an empirical analysis with a clear method of a full-fledged AI system, *Copycat* is performed. Different from many existing humanistic readings of computational systems, the analysis addresses both the rhetoric around, and operation of, the system. In other words, the method combines both a computer science analysis and a human interpretation of meanings. Compared with related works in software studies, the system under study is substantial in terms of its complexity and impact in the community. Furthermore, technical literature has been a relatively unexplored area for studying computer systems. However, as this dissertation has shown, the technical literature provides a rich resource for analysis. Finally, this approach can be generalized for other forms of code, not just AI systems.

Design Implications for Expressive Domains A new design strategy of agency play has been proposed for expressive domains. Drawn from social theories of agency, agency play intends to transform agency from a holy-grail in digital media design to an expressive resource that can convey meanings and express ideas. Agency play is also important from a social perspective because it encourages reflection of our human condition by allowing multiple understandings of agency in addition to “agency as free will.”

Memory, Reverie Machine *MRM* is designed to explore new design possibilities fostered by the new design theories of agency play and scale of intentionality. It offers a concrete example of the new design possibilities in an expressive domain by creating a character that oscillates between low levels of intentionality (an avatar) and high levels (a non-player character). Inspired by the aesthetics in Woolf's *Mrs. Dalloway*[218], *MRM*'s emphasis on the associative connections between events and memories and its loose story structure offers a unique narrative experience that is different from many planning-based interactive narrative works.

7.3 *Future Directions*

There are several aspects in this work that are subject to future exploration. This section provides a brief discussion of some of the promising future directions in both theoretical inquiries and technical/expressive production.

Empirical Study of User Interpretation This document has focused primarily on system authors' narration of system intentionality. Although some discussion has been devoted to users' interpretations an empirical user study is needed to support and expand this aspect of the AI hermeneutic network.

Software Hermeneutics The phenomenon of system intentionality is shared by a large range of digital and other technical artifacts. It is a broad area that will only grow in social significance and open up a number of possible research directions and applications in the domains of interactive narrative, gaming, interactive art installation, robotics, and human computer interaction (HCI).

Multicultural Perspectives of System Intentionality The history of computing has been narrated and studied mainly from a Euro-American vantage point. However, computing and software is not as culturally-neutral as many computer scientists

believe it to be. As Harrell puts forth in his article [79], there are other culturally grounded systems to pursue and even European/American grounded cultural foundations could be made explicit. The issue of the cultural roots of computing is of particular relevance to the framework of intentional systems, as the latter emphasizes the interpretive process of users, who inevitably bring their own cultural values and heritage. An important question is how would a different cultural point of view influence the narration and interpretation of system intentionality?

Extending *MRM* *MRM* is designed, in part, for inquiry into what is seen as especially promising avenues for research into agency play and scale of intentionality. Algorithmically, *MRM* can be extended to be more generative. For instance, the plot development is currently hard-coded and does not vary according to the user interaction. To enhance the manifestation of system intentionality and agency, I plan to explore various story generation techniques. A particularly interesting line of future research is the use of analogy to structure and modify the course of the story. For instance, events in the recalled memories can be transferred to the main story world through analogical mapping.

For literary artists and digital media arts practitioners, rich prose, genre experimentation, and interaction design will perhaps be seen as the ripest areas for future exploration. The current interface that requires the user to input keywords can be easily bolstered into, for instance, a graphical user interface. A user could also perform meaningful actions such as opening doors, washing dishes, setting visual scanners, and more through a visual gesture interface mechanism. Alternately, in a text-based system, user input could certainly be more complex than interactive fiction styled simple keywords, using for instance a natural language-based interface.

This dissertation celebrates the ever-changing technology and the new expressions

that it brings along. At heart, it is about humans. The proposed theory and design is meant to underscore the warmth of the human touch even in what seems to be the coldest machine piece or the most objective science/engineering practice. At its best, it is also about dreams and memories. We create these lifeless artifacts, digital or analog, and choose to believe that they share similar connection to the world as us as a way to pursue our dreams and embody our memories. At the centre of all above, is the intentionality, both of human and of machine.

APPENDIX A

SAMPLE DATA FROM THE TECHNICAL LITERATURE ON *COPYCAT*

Table 11: Use of Intentional Vocabulary from paper
“The Copycat Project: A Model of Mental Fluidity and
Analogy-making” [98].

Term	Context	Source
-	An earthquake takes place when subterranean structures are under sufficient pressure that something suddenly slips. Without the pressure, obviously, there would be no slippage. Analogous statement holds for pressures bringing about conceptual slippage: only under specific pressures will concepts slip into related ones.	[98, pp.207]
-	That is, the Copycat project is not about simulating analogy-making per se, but about simulating the very crux of human cognition: fluid concepts	[98, pp.208]
know	Just as the program knows the immediate neighbors of every letter in the alphabet, it also knows the successors and predecessors of small integers.	[98, pp.210]

know	Just as the program knows the immediate neighbors of every letter in the alphabet, it also knows the successors and predecessors of small integers.	[98, pp.210]
treat	Copycat can even treat small integers as it does letters - it can notice relationship between numbers, can group numbers together, map them onto each other, and so on.	[98, pp.210]
notice	Copycat can even treat small integers as it does letters - it can notice relationship between numbers, can group numbers together, map them onto each other, and so on.	[98, pp.210]
group	Copycat can even treat small integers as it does letters - it can notice relationship between numbers, can group numbers together, map them onto each other, and so on.	[98, pp.210]
map	Copycat can even treat small integers as it does letters - it can notice relationship between numbers, can group numbers together, map them onto each other, and so on.	[98, pp.210]
recognize	Thus, copycat is capable of recognizing not only that the structure	[98, pp.210]
resist	Copycat tends to resist bringing numbers into the picture, unless there seems to be some compelling reason to do so	[98, pp.210]

-	Finally, while humans tend to scan strings of roman letters from left to right, are much better at recognizing forwards alphabetical order than backwards alphabetical order, and have somewhat greater familiarity with the beginning of the alphabet than its middle or end, the Copycat program is completely free of these biases. This should not be regarded as a defect of the program, but a strength, because it keeps the project's focus away from the domain-specific and nongeneralizable details.	[98, pp.210]
-	In fact, the inspiration for the architecture comes in part from a computer model of low-level and high-level auditory perception: the Hearsay II speech-understanding project.	[98, pp.210]
-	Creative human thinkers manifest an exquisite selectivity of this sort - when they are faced with a novel situation, what bubbles up from their unconscious and pops to mind is typically a small set of concepts that "fit like a glove", without a host of extraneous and irrelevant concepts being consciously activated or considered.	[98, pp.211]
long-term memory	It (the Slipnet) can thought of, roughly, as Copycat's long-term memory	[98, pp.211]
-	The Workplace is the locus of perceptual activity	[98, pp.211]

-	The basic image for the Slipnet is that of a network of interrelated concepts.	[98, pp.212]
-	Nodes acquire varying levels of activation (which can be thought of as a measure of relevance to the situation at hand)	[98, pp.212]
-	The further away a given aspect of a situation is from direct perception, the more likely it is to be involved in what people consider to be the essence of the situation	[98, pp.212]
drive	We therefore built into the architecture a strong drive, if a deep aspect of a situation is perceived, to use it and to try to let it influence further perception of the situation	[98, pp.213]
propensity	there is a built-in propensity in the program to prefer slipping shallow concepts rather than deep concepts, when slippages have to be made	[98, pp.214]
prefer	there is a built-in propensity in the program to prefer slipping shallow concepts rather than deep concepts, when slippages have to be made	[98, pp.214]
mold	Slipnet is constantly altering its “shape” in attempting to mold itself increasingly accurately to fit the situation at hand	[98, pp.214]
-	It (copycat) aims at modeling cognitive-level behavior by simulating processes at a subcognitive but superneural level.	[98, pp.216]

-	The basic image for the Workspace is that of a busy construction site in which structures of many sizes and at many locations are being worked on simultaneously by independent crews, some occasionally being torn down to make way for new, hopefully better ones.	[98, pp.216]
-	this image comes essentially from the biological cell; the Workspace corresponds roughly to the cytoplasm of a cell, in which enzymes carrying out diverse tasks all throughout the cell's cytoplasm are the construction crews, and the structures built up are all sorts of hierarchically-structured biomolecules.	[98, pp.216]
unhappiness	is determined by the object's salience, which is a function of both the object's importance and unhappiness	[98, pp.217]
unhappy	An unhappy object is one that has few or no connections to the rest of the object in the Worldspace, and that thus seems to cry out for more attention	[98, pp.217]
cry out	An unhappy object is one that has few or no connections to the rest of the object in the Worldspace, and that thus seems to cry out for more attention	[98, pp.217]

-	the architecture honors the old motto “The squeaky wheel gets the oil”, even if only probabilistically so.	[98, pp.217]
-	In particular, people are clearly quicker to recognize two neighboring objects as identical than as being related in some abstract way. Thus the architecture has an intrinsic speed-bias in favor of sameness bonds: it tends to spot them and to construct them more quickly than it spots and constructs bonds representing other kinds of relationships.	[98, pp.217]
-	The fact that a and k are unrelated letters of the alphabet is simply ignored by most people. Copycat is constructed to behave similarly.	[98, pp.219]
-	all the aforementioned types of perceptual actions - scanning, bond-making, group-making, bridge-building, and so forth - take place in parallel.	[98, pp.219]
-	A useful image is that of highly coherent macroscopic structures (e.g. physical bridges) built by a colony of thousands of myopic ants or termites working semi-independently but nonetheless cooperatively. (The “ants” of Copycat - namely, codelets - will be described in the next subsection.)	-
-	the system is thus far more like a swarm of ants than like a rigid military hierarchy.	[98, pp.220]

-	We now turn to the description of Copycat’s “ants” [of codelets] and how they are biased.	[98, pp.220]
-	There are two types of codelets: scout codelets and effector codelets	[98, pp.221]
notice	For example, one scout codelet might notice that the adjacent r’s in mrrjjj are instances of the same letter	[98, pp.221]
desire	Top-down codelets represent specific pressures evoked by the specific situation at hand (e.g. the desire, in Problem 1 and 2, to look for more successor relations, once some have already been discovered).	[98, pp.222]
-	The mission of such a codelet is to scan the Workspace in search of instances of its spawning concept.	[98, pp.222]
-	The shifting population of codelets on the Coderack bears a close resemblance to the shifting enzyme population of a cell, which evolves in a sensitive way in response to the ever-changing makeup of the cell’s ectoplasm.	[98, pp.223]
feel	Whenever a codelet is given the chance to run, the various pressures for which it is a proxy make themselves slightly felt	[98, pp.224]
-	A useful image here is that of the course of play in a basketball game.	[98, pp.225]

-	<p>Much the same could be said for the Copycat: an outside observer is free, after a run is over, to “parse” the run in terms of specific, discrete processes, and to attempt to impose such a vocabulary on the system’s behavior; however, that parsing and labeling is not intrinsic to the system, and such interpretations are in no way unique or absolute, any more than in the basketball game. In other words, a long sequence of codelet actions can add up to what could be perceived, a posteriori and by an outsider; as a single coherent drive towards a particular goal, but that is the outsider’s subjective interpretation.</p>	[98, pp.226]
fingers of exploration	<p>The basic image is that of many “fingers of exploration” simultaneously feeling out various potential pathways at different speeds.</p>	[98, pp.226]
fight	<p>This entails a “fight” between the incumbent structure and the upstart</p>	[98, pp.226]
-	<p>This is how the system’s actual viewpoint develops with time</p>	[98, pp.226]
-	<p>Incidentally, this aspect of Copycat reflects the psychologically important fact that conscious experience is essentially unitary, although it is of course an outcome of many parallel unconscious processes.</p>	[98, pp.226]

-	A metaphor for the parallel terraced scan is provided by the image of a vast column of ants marching through a forest	[98, pp.226]
open-minded	the system loses this naive, open-minded quality; as indeed it ought to and usually ends up being “closed-minded”.	[98, pp.228]
close-minded	the system loses this naive, open-minded quality; as indeed it ought to and usually ends up being “closed-minded”.	[98, pp.228]
open-minded	At the start of a run, the system is open-minded, and for good reasons: it knows nothing about the situation it is facing	[98, pp.228]
knows	At the start of a run, the system is open-minded, and for good reasons: it knows nothing about the situation it is facing	[98, pp.228]
facing	At the start of a run, the system is open-minded, and for good reasons: it knows nothing about the situation it is facing	[98, pp.228]
-	Reflecting the system’s uncertain advances and retreats as it builds and destroys structures in its attempts to home in on the best way to look at a situation	[98, pp.228]
-	The codelet will thus tend to be more attracted to A than to B	[98, pp.229]
willing	Thus, the system itself controls the degree to which it is willing to take risks	[98, pp.229]

personality	and those data provided some of the most important insights into the program's "personality"	[98, pp.230]
-	The earlier image of the swiftly-changing panorama of a basketball game may help to make this necessary necction between asynchronous parallelism and randomness more intuitive	[98, pp.232]
-	except in the absurd sense of millions of monkeys plunking away on piano keyboards for trillions of years and coming up with "Blue Moon" once in a blue moon	[98, pp.232]
-	when there is a lot of fog ahead, and their purpose is precisely to get an evenly-distributed sense of what lies out there in the fog rather than simply plunging ahead blindly, at random	[98, pp.233]
-	The foggier things are, the more unbiased should be the scouting mission.	[98, pp.233]
deal	demonstrates how [the program] deals with these issues, how it responds to variations in pressues, and how it is able, starting from exactly the same state on each new problem, to fluidly adapt to a range of different situations.	[98, pp.234]
responds	demonstrates how [the program] deals with these issues, how it responds to variations in pressues, and how it is able, starting from exactly the same state on each new problem, to fluidly adapt to a range of different situations.	[98, pp.234]

adapt	demonstrates how [the program] deals with these issues, how it responds to variations in pressures, and how it is able, starting from exactly the same state on each new problem, to fluidly adapt to a range of different situations.	[98, pp.234]
-	the program assesses ijl. to be somewhat higher quality than ijd.	[98, pp.236]
understand- ing	that it had, in some sense, a reasonable “understanding” of what was going on in the problem.	[98, pp.236]
-	An alternative many people prefer is ljk..	[98, pp.238]
-	Time and cognitive resources being limited, it is vital to resist nonstandard ways of looking at situations without strong pressure to do so. You don’t check the street sign at the corner, every time you go outdoors, to reassure yourself that your street’s name hasn’t been changed.	[98, pp.240]
-	In a sense, answer wyz to Problem 6 seems like a miniature “conceptual revolution” or “paradigm shift”.	[98, pp.244]
like	The answer xyz, whose very high temperature of 74 indicates that the program did not “like” it at all,	[98, pp.246]
desperation	It takes the “desperation” caused by the z-snap to allow such strange ideas any chance at all	[98, pp.246]
-	the persistent high temperature - make the normally unappealing r-z bridge a bit more tempting	[98, pp.247]

-	the program tends to see sameness groups quite fast	[98, pp.249]
-	It would be disastrous if Copycat were willing to bring in unlikely notions.	[98, pp.249]
-	coupled with the system's unhappiness as it failure to.	[98, pp.249]
-	the successor relations among them might be spotted by bottom-up codelets.	[98, pp.249]
-	a small renegade codelet, totally unaware of the global momentum, has had some good luck.	[98, pp.255]
take calculated risk	Copycat opts for a middle way, in which it quite literally takes calculated risks all the time	[98, pp.256]
-	both source and target strings come quite quickly to be perceived as successor groups.	[98, pp.257]
-	encouraging a broader and more open-minded search	[98, pp.258]
-	the system's attachment to a viewpoint already established	[98, pp.258]
sniffed out	unusual avenues are more likely to at least be "sniffed out" a short ways.	[98, pp.259]
given the chance to	One such neighbor-concept is alphabetic-first, which is now briefly given the chance to show its relevance	[98, pp.259]
-	they almost cry out to be mapped onto each other - not because the system can anticipate the great insight that such a mapping will bring	[98, pp.259]

-	they almost cry out to be mapped onto each other - not because the system can anticipate the great insight that such a mapping will bring	[98, pp.259]
-	Once the system tries it out	[98, pp.259]
-	which means that the emerging new view will tend to get supported	[98, pp.260]
-	building a completely new view of the situation	[98, pp.261]
-	the process of arriving at answer wuz to be very similar, on an abstract level, to the process whereby a full-scale conceptual revolution takes place in science	[98, pp.261]
-	Copycat was forced to look for solutions	[98, pp.261]
-	Copycat was lucky enough to come across the double reversal.	[98, pp.262]
-	the system is focusing its attention on just a small set of concepts.	[98, pp.266]

REFERENCES

- [1] ABU-LUGHOD, L., “The romance of resistance: Tracing transformations of power through bedouin women,” *American Ethnologist*, vol. 17, no. 1, pp. 41–55, 1990.
- [2] AGRE, P. E., *Computation and Human Experience*. Cambridge, U.K.: Cambridge University Press, 1997.
- [3] AGRE, P. E., “Toward a critical technical practice: Lessons learned in trying to reform ai,” in *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide* (BOWKER, G. C., STAR, S. L., TURNER, W., GASSER, L., and BOWKER, G., eds.), pp. 131–158, Lawrence Erlbaum Associates, 1997.
- [4] AHEARN, L. M., “Language and agency,” *Annual Review of Anthropology*, vol. 30, pp. 109–137, 2001.
- [5] ALLEN, W., “Another woman (film),” 1988.
- [6] BACCHUS, F., KABANZA, F., and SHERBROOKE, U. D., “Using temporal logics to express search control knowledge for planning,” *Artificial Intelligence*, vol. 116, p. 2000, 1999.
- [7] BARTHES, R., *Mythologies*. New York: Noonday Press, 1973.
- [8] BARTHES, R., *S/Z*. New York: Hill and Wang, 1974.
- [9] BARTHES, R., *Image-Music-Text*. New York: Hill and Wang, 1977.
- [10] BATES, J., “The role of emotion in believable agents,” *Communications of the ACM*, vol. 37, pp. 122–125, 1994.

- [11] BEER, R. D., “Dynamical approaches to cognitive science,” *Trends in Cognitive Sciences*, vol. 4, no. 3, pp. 91–99, 2000.
- [12] BOBROW, D. G., “Natural language input for a computer problem solving system,” tech. rep., Massachusetts Institute of Technology, Cambridge, MA, USA, 1964.
- [13] BODEN, M. A., *The Creative Mind*. London: Abacus, 1992.
- [14] BOGOST, I., *Unit Operations*. Cambridge, MA: The MIT Press, 2006.
- [15] BOLAND, R. J., “Information systems use as a hermeneutic process,” in *Information Systems Research: Contemporary Approaches Emergent Traditions* (NISSEN, H.-E., KLEIN, H. K., and HIRSCHHEIM, R., eds.), pp. 439–458, New York: North-Holland, 1991.
- [16] BRACHMAN, R. and LEVESQUE, H., *Knowledge Representation and Reasoning*. Morgan Kaufmann, 2004.
- [17] BRAITENBERG, V., *Vehicles: Experiments in synthetic psychology*. Cambridge: MIT Press, 1984.
- [18] BRENTANO, F., *Psychology from an Empirical Standpoint*. London: Routledge & Kegan Paul, 1874.
- [19] BRINGSJORD, S. and FERRUCCI, D. A., *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine*. Hillsdale, NJ: Lawrence Erlbaum, 2000.
- [20] BROOKS, R. A., “Elephants don’t play chess,” *Robotics and Autonomous Systems*, no. 6, pp. 3–15, 1990.
- [21] BROOKS, R. A., “Intelligence without representation,” *Artificial Intelligence*, vol. 47, 1991.

- [22] BUSH, V., “As we may think,” *Atlantic Monthly*, pp. 101–108, July 1945 1945.
- [23] CALLON, M., “Some elements of a sociology of translation: Domestication of the scallops and the fishermen of st brieuc bay,” in *Power, Action and Belief: A New Sociology of Knowledge* (LAW, J., ed.), London: Routledge & Kegan Paul, 1986.
- [24] CARBONELL, J., MICHALSKI, R., , and MITCHELL, T., *Machine Learning, an Artificial Intelligence Approach*. Tioga Publishing, 1983.
- [25] CLARK, A., *Natural-born cyborgs : minds, technologies, and the future of human intelligence*. New York: Oxford University Press, 2003.
- [26] COLE, D., “The chinese room argument,” in *The Stanford Encyclopedia of Philosophy* (ZALTA, E. N., ed.), <http://plato.stanford.edu/archives/fall2008/entries/chinese-room/>, Fall 2008.
- [27] COLTON, S., “Creativity versus the perception of creativity in computational systems,” in *Proceedings of the AAAI 2008 Spring Symposium in Creative Intelligent Systems*, (Stanford), AAAI Press, 2008.
- [28] COPE, D., *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, MA: The MIT Press, 2004.
- [29] CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L., and STEIN, C., *Introduction to Algorithms*. Cambridge: MIT Press, 2001.
- [30] COX, G., MCLEAN, A., and WARD, A., “The aesthetics of generative code,” <http://www.generative.net/papers/aesthetics/>, 2001.
- [31] CRAMER, F., “Concepts, notations, software, art,” in *Seminar for Allegemeine und Vergleichende Literaturwissenschaft*, 2002.

- [32] CREVIER, D., *AI: The Tumultuous Search for Artificial Intelligence*. New York: Basic Books, 1993.
- [33] DAVIDSON, D., “Agency,” in *Essays on Actions and Events: Philosophical Essays*, pp. 43–62, Oxford, UK: Oxford University Press, 2001.
- [34] DAWKINS, R., *The Selfish Gene*. London: Paladin, 1976.
- [35] DENNETT, D. C., “True believers: the intentional strategy and why it works,” in *Mind Design II: Philosophy Psychology Artificial Intelligence* (HAUGELAND, J., ed.), pp. 57–79, MIT Press, 1981.
- [36] DENNETT, D. C., *Kinds of Minds*. New York: Basic Books, 1996.
- [37] DENNETT, D. C. and HAUGELAND, J., “Intentionality,” in *The Oxford Companion to the Mind* (GREGORY, R. L., ed.), Oxford: Oxford University Press, 1987.
- [38] DENNETT, D. C., *The Intentional Stance*. Cambridge: MIT Press, 1987.
- [39] DILTHEY, W., *Selected Writings*. Cambridge: Cambridge University Press, 1976.
- [40] DIX, A., FINLAY, J., ABOWD, G., and BEALE, R., *Human-Computer Interaction*. Edinburgh Gate, England: Prentice Hall, 2003.
- [41] DOUGLAS, R. L., “Formalizing an african-american aesthetic,” *New Art Examiner*, p. 1824, 1991.
- [42] DREYFUS, H. L., *What Computers Can’t Do: A Critique of Artificial Reason*. San Francisc: Freeman, 1972.
- [43] DRUMMOND, C., “Machine learning as an experimental science (revisited),” in *Proceedings of the Twenty-First National Conference on Artificial Intelligence*:

- Workshop on Evaluation Methods for Machine Learning*, pp. 1–5, AAAI Press, 2006.
- [44] DUDA, R., GASCHING, J., and HART, P., “Model design in the prospector consultant system for mineral exploration,” in *Expert Systems in the Micro-Electronic Age* (MICHIE, D., ed.), pp. 153–167, Edinburgh: Edinburgh University Press, 1979.
 - [45] EDWARDS, P. N., *The closed world : computers and the politics of discourse in Cold War America*. Cambridge, MA: MIT Press, 1996.
 - [46] ELLIOTT, C. and BRZEZINSKI, J., “Autonomous agents as synthetic characters,” *AI Magazine*, vol. 19, no. 2, pp. 13–30, 1998.
 - [47] ELMAN, J. L., “Connectionism, artificial life, and dynamical systems,” in *A Companion to Cognitive Science* (BECHTEL, W. and GRAHAM, G., eds.), pp. 488–505, Oxford: Basil Blackwood, 1998.
 - [48] ENGELBART, D. C., “Augmenting human intellect: A conceptual framework,” tech. rep., Stanford Research Institute, 1962.
 - [49] EVANS, V., BERGEN, B. K., and ZINKEN, J., “The cognitive linguistics enterprise: An overview,” in *The Cognitive Linguistics Reader* (EVANS, V., BERGEN, B. K., and ZINKEN, J., eds.), Advances in Cognitive Linguistics, London, U.K.: Equinox Press, 2006.
 - [50] FALKENHAINER, B., FORBUS, K. D., and GENTNER, D., “The structure-mapping engine: Algorithm and examples,” *Artificial Intelligence*, vol. 41, pp. 1–63, 1989.
 - [51] FAUCONNIER, G., *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: MIT Press/Bradford Books, 1985.

- [52] FAUCONNIER, G., “Methods and generalizations,” in *Cognitive Linguistics: Foundations, Scope, and Methodology* (JANSSEN, T. and REDEKER, G., eds.), pp. 95–128, Mouton de Gruyter: The Hague, 1999.
- [53] FAUCONNIER, G., “Conceptual blending and analogy,” in *The Analogical Mind: Perspectives from Cognitive Science* (GENTNER, D., HOLYOAK, K. J., and KOKINO, B. N., eds.), Cambridge, MA: MIT Press, 2001.
- [54] FAUCONNIER, G. and TURNER, M., *The Way We Think: Conceptual Blending and the Mind’s Hidden Complexities*. New York, NY: Basic Books, 2002.
- [55] FERNANDEZ, J. W., “Introduction: Confluents of inquiry,” in *Beyond Metaphor: The Theory of Tropes in Anthropology* (FERNANDEZ, J. W., ed.), pp. 3–13, Stanford University Press, 1991.
- [56] FLOR, N. V. and HUTCHINS, E., “Analyzing distributed cognition in software teams: A case study of team programming during perfective software maintenance,” in *Proceedings of the Fourth Annual Workshop on Empirical Studies of Programmers*, pp. 36–59, Norwood, NJ: Ablex Publishing, 1991.
- [57] FORBUS, K. D., “Exploring analogy in the large,” in *The analogical mind* (GENTNER, D., HOLYOAK, K. J., and KOKINOV, B. N., eds.), pp. 23–58, Cambridge: MIT Press, 2000.
- [58] FORD, K. M., GLYMOUR, C., and HAYES, P. J., “On the other hand - cognitive prostheses,” *AI Magazine*, vol. 18, no. 3, p. 104, 1997.
- [59] FOUCAULT, M., *Archeology of Knowledge*. New York: Pantheon, 1972.
- [60] FOUCAULT, M., *Discipline and Punish*. New York: Pantheon, 1977.
- [61] FOUCAULT, M., *The History of Sexuality, Vol. I: An Introduction*, vol. 1. New York: Pantheon, 1978.

- [62] FRANK, M., *The Subject and the Text: Essays on literary theory and philosophy*. Cambridge: Cambridge University Press, 1989.
- [63] FRANKLIN, S., “Ida, a conscious artifact?,” *Journal of Consciousness Studies*, vol. 10, pp. 47–66, 2003.
- [64] FRENCH, R. M., *The subtlety of sameness: a theory and computer model of analogy-making*. Cambridge: MIT Press, 1995.
- [65] FULLER, M., *Behind the Blip: Essays on the Culture of Software*. New York: Autonomedia, 2003.
- [66] FULLER, M., *Software Studies: A Lexicon*. Cambridge: MIT Press, 2008.
- [67] GADAMER, H.-G., *Reason in the Age of Science*. Cambridge, MA: MIT Press, 1983.
- [68] GENETTE, G., *Narrative Discourse: An Essay in Method*. Ithaca, NY: Cornell University Press, 1980.
- [69] GIBBS JR., R. W., “Making good psychology out of blending theory,” *Cognitive Linguistics*, vol. 11.3/4, pp. 347–358, 2000.
- [70] GIDDENS, A., *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis*. Berkeley: University of California Press, 1979.
- [71] GOGUEN, J., “An introduction to algebraic semiotics, with applications to user interface design,” in *Computation for Metaphors, Analogy, and Agents* (NEHANIV, C., ed.), (Yakamatsu, Japan), 1998.
- [72] GOGUEN, J. and HARRELL, D. F., “Style as a choice of blending principles,” in *Style and Meaning in Language, Art, Music and Design, Proceedings of the 2004 AAAI Fall Symposium* (DUBNOV, S. and ARGAMON, S., eds.), vol. FS-04-07, (Washington D.C.), AAAI Press, 2004.

- [73] GRADY, J. E., OAKLEY, T., and COULSON, S., “Blending and metaphor,” in *Metaphor in Cognitive Linguistics* (STEEN, G. and GIBBS, R., eds.), Amsterdam: John Benjamins, 1999.
- [74] GREENBERG, C., “Modernist painting,” 1960.
- [75] GUNNING, T., *D.W. Griffith and the Origins of American Narrative Film: THE EARLY YEARS AT BIOGRAPH*. Champaign, IL: University of Illinois Press, 1993.
- [76] HARAWAY, D., *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*. Routledge, 1991.
- [77] HARRELL, D. F., “Shades of computational evocation and meaning: The griot system and improvisational poetry generation,” in *6th Digital Arts and Culture Conference*, (Copenhagen, Denmark), pp. 133–143, 2005.
- [78] HARRELL, D. F., “Walking blues changes undersea: Imaginative narrative in interactive poetry generation with the griot system,” in *AAAI 2006 Workshop in Computational Aesthetics: Artificial Intelligence Approaches to Happiness and Beauty*, (Boston, MA), AAAI Press, 2006.
- [79] HARRELL, D. F., “Cultural roots for computing: The case of african diasporic orature and computational narrative in the griot system,” *Fiberculture*, no. 11, 2007.
- [80] HARRELL, D. F., “Griot’s tales of haints and seraphs: A computational narrative generation system,” in *Second Person: Role-Playing and Story in Games and Playable Media* (HARRIGAN, N. W.-F. and PATRICK, eds.), Cambridge, MA: The MIT Press, 2007.

- [81] HARRELL, D. F., *Theory and Technology for Computational Narrative: An Approach to Generative and Interactive Narrative with Bases in Algebraic Semiotics and Cognitive Linguistics*. Dissertation, University of California, San Diego, 2007.
- [82] HARRELL, D. F., “Algebra of identity: Skin of wind, skin of streams, skin of shadows, skin of vapor,” in *Critical Digital Studies* (KROKER, A. and KROKER, M., eds.), Toronto: University of Toronto Press, 2008.
- [83] HARRELL, D. F., “Toward a theory of phantasmal media:an imaginative cognition and computation-based approach to digital media,” *CTheory*, 2009 (forthcoming).
- [84] HARRELL, D. F. and ZHU, J., “Agency play: Dimensions of agency for interactive narrative design,” in *Proceedings of AAAI 2009 Spring Symposium on Interactive Narrative Technologies II*, (Stanford, CA), p. in press, AAAI Press, 2009.
- [85] HAUGELAND, J., *Artificial intelligence: the very idea*. Cambridge: MIT Press, 1985.
- [86] HAWKINS, J. and BLAKESLEE, S., *On Intelligence*. Times Books, 2004.
- [87] HAYLES, N. K., *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press, 1999.
- [88] HEIDEGGER, M., *Being and Time*. San Francisco: Harper, 1962.
- [89] HEIDEGGER, M., “The question concerning technology,” in *Basic writings from Being and time (1927) to The task of thinking (1964)*, pp. xvi, 397 p., New York ; Toronto: Harper & Row, 1st ed., 1977.

- [90] HEIDER, F. and SIMMEL, M.-A., "An experimental study of apparent behavior," *American Journal of Psychology*, vol. 57, pp. 243–59, 1944.
- [91] HELMERT, M., *Understanding Planning Tasks: Domain Complexity and Heuristic Decomposition*. Berlin: Springer, 1998.
- [92] HIRSCH, E. D., *Validity in Interpretation*. New Haven and London: Yale University Press, 1967.
- [93] HOFSTADTER, D. R., *Gödel, Escher, Bach : an eternal golden braid*. New York: Basic Books, 1979.
- [94] HOFSTADTER, D. R., "The copycat project: An experiment in nondeterminism and creative analogies," Tech. Rep. AI Memo No. 755, AI Laboratory, MIT, 1984.
- [95] HOFSTADTER, D. R., *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic books, 1995.
- [96] HOFSTADTER, D. R., "Staring emmy straight in the eye - and doing my best not to flinch," in *Creativity, Cognition and Knowledge* (DARTNALL, T., ed.), pp. 67–100, Westport, CT: Praeger, 2002.
- [97] HOFSTADTER, D. R. and MITCHELL, M., "The copycat project: A model of mental fluidity and analogy-making," in *Advances in connectionist and neural computation theory: Analogical connections* (BARNDEN, J. A. and HOLYOAK, K. J., eds.), vol. 2, pp. 31 – 112, Norwood, NJ: Ablex, 1994.
- [98] HOFSTADTER, D. R. and MITCHELL, M., "The copycat project: A model of mental fluidity and analogy-making," in *Fluid concepts & creative analogies* (HOFSTADTER, D. R., ed.), pp. 205–268, New York: Basic Books, 1995.
- [99] HOLMES, B., "Requiem for the soul," *New Scientist*, no. 2094, 1997.

- [100] HOLQUIST, M., *Dialogism: Bakhtin and His World*. Routledge, 2 ed., 1990.
- [101] HOLYOAK, K. J. and THAGARD, P., "Analogical mapping by constraint satisfaction," *Cognitive Science*, vol. 13, pp. 295–355, 1989.
- [102] HSU, F.-H., *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*. Princeton, NJ: Princeton University Press, 2002.
- [103] HUMPHREY, R., *Stream of Consciousness in the Modern Novel*. Perspectives in Criticism, Berkeley and Los Angeles: University of California Press, 1954.
- [104] HUSSERL, E., *Logical Investigations*. Routledge, 1970.
- [105] HUTCHINS, E., *Cognition in the Wild*. Cambridge: MIT Press, 1994.
- [106] HUTCHINS, E., "How a cockpit remembers its speed.," *Cognitive Science*, vol. 19, pp. 265–288, 1995.
- [107] JAMES, W., *The Principles of Psychology*. New York: Henry Holt And Company, 1890.
- [108] JUUL, J., "Games telling stories?," *Game Studies*, vol. 1, no. 1, 2001.
- [109] KAY, A. and A, G., "Personal dynamic media," *IEEE Computer*, vol. 10, pp. 31–41, 1977.
- [110] KELLY, K., "By analogy: A talk with the most remarkable researcher in artificial intelligence today, douglas hofstadter, the author of gdel, escher, bach," *Wired Magazine*, vol. November 1995, 1995.
- [111] KING, W. J., "Anthropomorphic agents: Friend, foe, or folly," tech. rep., University of Washington, 1995.
- [112] KIRSCHENBAUM, M. G., "Virtuality and vrml: Software studies after manovich," *Electronic Book Review*, 2003.

- [113] KITTLER, F., “There is no software,” 1995.
- [114] KUBRICK, S., “2001: A space odyssey (film),” 1968.
- [115] KUUTTI, K., “Activity theory and its applications to information systems research and development,” in *Information System Research* (NISSEN, H.-E., KLEIN, H. K., and HIRSCHHEIM, R., eds.), pp. 529–549, Amsterdam: Elsevier Science Publishers, 1991.
- [116] KUUTTI, K., “Activity theory as a potential framework for human-computer interaction research,” in *Context and Consciousness* (NARDI, B., ed.), pp. 17–44, Cambridge: MIT Press, 1995.
- [117] LAKOFF, G., “The contemporary theory of metaphor,” in *Metaphor and Thought (2nd ed)* (ORTONY, A., ed.), pp. 202–251, Cambridge University Press, 1993.
- [118] LAKOFF, G. and JOHNSON, M., *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.
- [119] LAKOFF, G. and JOHNSON, M., *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Cambridge, MA: MIT Press, 1999.
- [120] LANGLEY, P., “Machine learning as an experimental science,” *Mach. Learn.*, vol. 3, no. 1, pp. 5–8, 1988.
- [121] LATOUR, B., “Mixing humans and nonhumans together: The sociology of a door-closer,” *Social Problems*, vol. 35, no. 3, pp. 298–310, 1988.
- [122] LATOUR, B., *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge: Harvard University Press, 1988.
- [123] LATOUR, B., *We Have Never Been Modern*. Cambridge: Harvard University Press, 1993.

- [124] LATOUR, B., *Aramis, or the Love of Technology*. Cambridge: Harvard University Press, 1996.
- [125] LENAT, D. B., “Cyc: a large-scale investment in knowledge infrastructure,” *Commun. ACM*, vol. 38, no. 11, pp. 33–38, 1995.
- [126] LENAT, D. B. and BROWN, J. S., “Why am and eurisko appear to work,” *Artificial Intelligence*, vol. 23, no. 3, pp. 269–294, 1984.
- [127] LEONT’EV, A., “The problem of activity in psychology,” *Soviet Psychology*, vol. 13, no. 2, pp. 4–33, 1974.
- [128] LEVESQUE, H. J. and BRACHMAN, R. J., “Expressiveness and tractability in knowledge representation and reasoning,” *Computational Intelligence*, vol. 3, no. 2, pp. 78–93, 1987.
- [129] LEWIS, G. E., “Too many notes: Computers, complexity and culture in voyager,” *Leonardo Music Journal*, vol. 10, pp. 33–39, 2000.
- [130] LICKLIDER, J., “Man-computer symbiosis,” *IRE Transactions on Human Factors in Electronics*, vol. HFE-1, pp. 4–11, 1960.
- [131] LIU, H. and SINGH, P., “Conceptnet—a practical commonsense reasoning toolkit,” *BT Technology Journal*, vol. 22, no. 4, pp. 211–226, 2004.
- [132] LUCASARTS, “Star wars: Knights of the old republic (xbox),” 2003.
- [133] MALLERY, J. C., HURWITZ, R., and DUFFY, G., “Hermeneutics: from textual explication to computer understanding?,” in *The Encyclopedia of Artificial Intelligence* (SHAPIRO, S. C., ed.), New York: John Wiley & Sons, 1987.
- [134] MANOVICH, L., *The Language of New Media*. Cambridge: MIT Press, 2001.

- [135] MATEAS, M., “Expressive ai: A hybrid art and science practice,” *Leonardo*, vol. 34, no. 2, pp. 147–153, 2001.
- [136] MATEAS, M., *Interactive Drama, Art, and Artificial Intelligence*. PhD thesis, CMU, 2002.
- [137] MATEAS, M., “Reading hal: Representation and artificial intelligence,” in *Stanley Kubrick’s 2001: A Space Odyssey: New Essays* (KOLKER, R., ed.), Oxford University Press, 2006.
- [138] MATEAS, M. and STERN, A., “A behavior language for story-based believable agents,” *IEEE Intelligent Systems*, vol. 17, no. 4, pp. 39–47, 2002.
- [139] MCCULLOCH, W. S. and PITTS, W., “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943.
- [140] MCDERMOTT, D., “Artificial intelligence meets natural stupidity,” *SIGART Bull*, no. 57, pp. 4 – 9, 1976.
- [141] MCLUHAN, M., *Understanding Media: The Extensions of Man*. New York: McGraw Hill, 1964.
- [142] MEEHAN, J., *The Metanovel: Writing Stories by Computer*. Ph.d., Yale University, 1976.
- [143] MINSKY, M., “A framework for representing knowledge,” tech. rep., Cambridge, MA, USA, 1974.
- [144] MINSKY, M., *The Society of Mind*. New York: Simon and Schuster, 1986.
- [145] MINSKY, M., “Thoughts about artificial intelligence,” in *The Age of Intelligent Machines* (KURZWEIL, R., ed.), The MIT Press, 1992.

- [146] MITCHELL, M., *COPYCAT: A Computer Model of High-level Perception and Conceptual Slippage in Analogy-making*. PhD thesis, University of Michigan, 1990.
- [147] MITCHELL, M. and HOFSTADTER, D. R., “Perspectives on copycat: Comparisons with recent work,” in *Fluid Concepts & Creative Analogies* (HOFSTADTER, D. R., ed.), pp. 275–299, New York: Basic Books, 1995.
- [148] MITCHELL, T., *Machine Learning*. McGraw-Hill Education, 1997.
- [149] MONTFORT, N., *GENERATING NARRATIVE VARIATION IN INTERACTIVE FICTION*. PhD thesis, University of Pennsylvania, 2007.
- [150] MONTFORT, N., “Ordering events in interactive fiction narratives on intelligent narrative technologies,” in *AAAI 2007 Fall Symposium on Intelligent Narrative Technologies*, pp. 87–94, AAAI Press, 2007.
- [151] MORAVEC, H. P., *Mind children : the future of robot and human intelligence*. Cambridge, Mass.: Harvard University Press, 1988.
- [152] MUELLER, E. T., *Daydreaming in Humans and Machines: A Computer Model of the Stream of Thought*. Norwood, NJ: Ablex Publishing Corporation, 1990.
- [153] MUELLER, E. T. and DYER, M. G., “Daydreaming in humans and computers,” pp. 278–280, 1985.
- [154] MURRAY, J. H., *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge: The MIT Press, 1998.
- [155] NARDI, B., “Activity theory and human-computer interaction,” in *Context and Consciousness* (NARDI, B., ed.), pp. 4–8, Cambridge: MIT Press, 1996.

- [156] NARDI, B., “Studying context: a comparison of activity theory, situated action models, and distributed cognition,” in *Context and Consciousness* (NARDI, B., ed.), pp. 35–52, Cambridge: MIT Press, 1996.
- [157] NELSON, T., *Computer Lib: Dream Machines*. Redmond: Tempus Books of Microsoft Press, 1987.
- [158] NEWELL, A., “The knowledge level,” *AI Magazine*, vol. 2, no. 2, pp. 1–33, 1981.
- [159] NEWELL, A., SHAW, J. C., and SIMON, H. A., “Report on a general problem-solving program,” in *Proceedings of the International Conference on Information Processing*, pp. 256–264, 1959.
- [160] NEWELL, A. and SIMON, H. A., “Computer science as empirical inquiry: symbols and search (1976),” in *Perspectives on the computer revolution*, pp. 109–133, Norwood, NJ: Ablex Publishing Corp., 1989.
- [161] NORMAN, D., *The Psychology of Everyday Things*. New York: Basic Books, 1988.
- [162] NORVIG, P., *Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp*. San Francisco: Morgan Kaufmann, 1991.
- [163] ORTNER, S. B., *Anthropology and Social Theory: Culture, Power, and the Acting Subject*. Durham, NC: Duke University Press, 2006.
- [164] PALMER, R. E., *Hermeneutics: Interpretation theory in Schleiermacher, Dilthey, Heidegger and Gadamer*. Evanston, IL: Northwestern University Press, 1969.
- [165] PARKER, J., “George lewis,” *BOMB: A quarterly arts & culture magazine since 1981*, Fall 2005 2005.

- [166] PENNY, S., “Representation, enaction, and the ethics of simulation,” in *First Person* (WARDROP-FRUIN, N. and HARRIGAN, P., eds.), Cambridge, MA: The MIT Press, 2004.
- [167] PÉREZ Y PÉREZ, R., *MEXICA: a Computer Model of Creativity in writing*. PhD thesis, 1999.
- [168] PÉREZ Y PÉREZ, R. and ALISEDA, A., “The role of abduction in automatic storytelling,” in *Proceedings of the AAAI workshop in Computational Aesthetics*, pp. 53–60, AAAI Press, 2006.
- [169] PÉREZ Y PÉREZ, R. and SHARPLES, M., “Three computer-based models of storytelling: Brutus , minstrel and mexica,” *Knowledge-Based Systems*, vol. 17, pp. 15–29, 2004.
- [170] PÉREZ Y PÉREZ, R., SOSA, R., and LEMAITRE, C., “A computer model for visual-daydreaming,” in *AAAI Fall Symposium on Intelligent Narrative Technologies*, (Washington D.C.), AAAI Press, 2007.
- [171] PERLIN, K. and GOLDBERG, A., “Improv: a system for scripting interactive actors in virtual worlds,” *Computer Graphics*, vol. 29, no. 3, 1996.
- [172] PIAGET, J., “The concept of life,” in *The Child’s Conceptual of the World* (PIAGET, J., ed.), pp. 194–206, Totowa, NJ: Littlefield, Adams, 1975.
- [173] PICKERING, A., *The Mangle of Practice*. University of Chicago Press, 1995.
- [174] RAMBERG, B. and GJESDAL, K., “Hermeneutics,” in *The Stanford Encyclopedia of Philosophy* (ZALTA, E. N., ed.), <http://plato.stanford.edu/archives/spr2009/entries/hermeneutics/>, Spring 2009.

- [175] RAY, T. S., “An evolutionary approach to synthetic biology: Zen and teh art of creating life,” *Artificial Life*, vol. 1, 1994.
- [176] RICOEUR, P., *Hermeneutics and the human sciences*. Cambridge: Cambridge University Press, 1981.
- [177] RIEDL, M., *Narrative Generation: Balancing Plot and Character*. PhD thesis, 2004.
- [178] RODDENBERRY, G., “Star trek: The next generation (tv series),” 1987-1994.
- [179] ROSCH, E., MERVIS, C., GRAY, W., JOHNSON, D., and BOYES-BRAEM, P., “Basic objects in natural categories,” *Cognitive Psychology*, vol. 8, no. 3, pp. 382–439, 1976.
- [180] RUSSELL, S. and NORVIG, P., *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd ed., 2002.
- [181] RYAN, M.-L., “Beyond myth and metaphor: The case of narrative in digital media,” *Game Studies*, vol. 1, no. 1, 2001.
- [182] RYCHTYCKYJ, N. and TURSKI, A., “Reasons for success (and failure) in the development and deployment of ai systems,” in *AAAI 2008 workshop on What Went Wrong and Why*, 2008.
- [183] SANTA ANA, O., *Brown Tide Rising: Metaphors of Latinos in Comtemporary American Public Discourse*. Austin: University of Texas Press, 2002.
- [184] SAUSSURE, F. D., *Course in General Linguistics*. the Philosophical Library, Inc., 1959.
- [185] SCHANK, R., *Dynamic memory: A theory of reminding and learning in computers and people*. New York: Cambridge University Press, 1983.

- [186] SCHLEIERMACHER, F., *Hermeneutik*. Heidelberg: Carl Winter Universitätsverlag, 1959.
- [187] SCOTT, R., “Blade runner,” 1982.
- [188] SEARLE, J., “Minds, brains, and programs,” *Behavioral and Brain Sciences*, vol. 3, pp. 417–424, 1980.
- [189] SEARLE, J., *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press, 1983.
- [190] SEEL, N., *Agent Theories and Architectures*. PhD thesis, Surrey University, 1989.
- [191] SEGA, “Shenmue,” 2000.
- [192] SEGAL, J. M., *Agency and Alienation: A Theory of Human Presence*. Lanham MD: Rowman & Littlefield, 1991.
- [193] SENGERS, P., *Anti-Boxology: Agent Design in Cultural Context*. PhD thesis, Carnegie Mellon University, 1998.
- [194] SHNEIDERMAN, B., “Direct manipulation: a step beyond programming languages,” *IEEE Computer*, vol. 16, no. 8, pp. 57–69, 1983.
- [195] SHNEIDERMAN, B. and MAES, P., “Direct manipulation vs interface agents: Excerpts from debates at iui 97 and chi 97,” *Interactions*, no. November + December, pp. 42–61, 1997.
- [196] SHOHAM, Y., “Agent-oriented programming,” in *Readings in Agents* (HUHNS, M. N. and SINGH, M. P., eds.), pp. 328–348, San Francisco: Morgan Kaufmann, 1991.

- [197] SHORTLIFFE, E., *Computer Based Medical Consultations: MYCIN*. American Elsevier, 1976.
- [198] SONY COMPUTER ENTERTAINMENT, “Shadow of the colossus (playstation 2),” 2005.
- [199] SPIELBERG, S., “Artificial intelligence: Ai,” 2001.
- [200] STURKEN, M. and CARTWRIGHT, L., *Practices of Looking: An Introduction to Visual Culture*. New York: Oxford University Press, 2001.
- [201] SUNG, J.-Y., GUO, L., GRINTER, R. E., and CHRISTENSEN, H. I., “my roomba is rambo: Intimate home appliances,” in *Proceedings of UbiComp 2007*, pp. 145–162, Springer, 2007.
- [202] TURING, A., “Computing machinery and intelligence,” *Mind*, vol. 59, pp. 433–460, 1950.
- [203] TURKLE, S., *The Second Self: Computers and the Human Spirit*. New York: Simon and Schuster, 1984.
- [204] TURKLE, S., *Life on the Screen: Identity in the Age of the Internet*. MIT Press, 1995.
- [205] TURNER, M., *The Literary Mind: The Origins of Thought and Language*. New York; Oxford: Oxford UP, 1996.
- [206] TURNER, M., “Double-scope stories,” in *Narrative Theory and the Cognitive Sciences* (HERMAN, D., ed.), Stanford, CA: CSLI Publications, 2003.
- [207] TURNER, S. R., *Minstrel: a computer model of creativity and storytelling*. PhD thesis, University of California at Los Angeles, Los Angeles, CA, USA, 1993.

- [208] VIRKLER, H. A., *Hermeneutics: Principles and Processes of Biblical Interpretation*. Grand Rapids, MI: Baker Book House, 1981.
- [209] WARDRIP-FRUIN, N., “The ”tale-spin” effect: Toward an acknowledgement of process in digital literature,” in *Proceedings of perthDAC 2007* (HUTCHISON, A., ed.), pp. 394–403, 2007.
- [210] WARDRIP-FRUIN, N., *Expressive Processing: Digital Fictions, Computer Games, and Software Studies*. Cambridge: MIT Press, forthcoming.
- [211] WEIZENBAUM, J., “Eliza – a computer program for the study of natural language communication between man and machine,” *Communications of the ACM*, vol. 9, pp. 46–45, 1966.
- [212] WEIZENBAUM, J., *Computer Power and Human reason*. New York: Freeman, 1976.
- [213] WIENER, N., *The Human Use of Human Beings : Cybernetics and Society*. Cambridge MA: Houghton-Mifflin, 1950.
- [214] WILLIAMS, R., *Television: Technology and Cultural Form*. New York: Schocken Books, 1975.
- [215] WINNER, L., “Do artifacts have politics?,” in *The whale and the reactor: a search for limits in an age of high technology* (WINNER, L., ed.), pp. 19–39, Chicago: University of Chicago Press, 1986.
- [216] WINOGRAD, T., *Understanding Natural Language*. New York: Academic Press, 1972.
- [217] WINOGRAD, T. and FLORES, F., *Understanding Computers and Cognition*. Norwood, NJ: Ablex Publishing Corporation, 1986.

- [218] WOOLF, V., *Mrs. Dalloway*. Harcourt, 2002 (1925).
- [219] YOUNG-HAE CHANG HEAVY INDUSTRIES, “Dakota,” 2002.
- [220] ZHU, J. and HARRELL, D. F., “Daydreaming with intention: Scalable blending-based imagining and agency in generative interactive narrative,” in *AAAI 2008 Spring Symposium on Creative Intelligent Systems*, (Stanford, CA), pp. 156–162, AAAI Press, 2008.
- [221] ZHU, J. and HARRELL, D. F., “Narrating artificial daydreams, memories, reveries: Toward scalable intentionality in expressive artificial intelligence practice. presented at the electronic literature organization conference (elo).,” 2008.